

YASMÍN ROMINA MONTENEGRO MALUENDA

**PROPRIEDADES DO ALGORITMO
LMS OPERANDO EM PRECISÃO
FINITA**

**FLORIANÓPOLIS
2005**

**UNIVERSIDADE FEDERAL DE SANTA
CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA**

**PROPRIEDADES DO ALGORITMO
LMS OPERANDO EM PRECISÃO
FINITA**

Dissertação submetida à
Universidade Federal de Santa Catarina
como requisito parcial à obtenção do grau de
Mestre em Engenharia Elétrica

YASMÍN ROMINA MONTENEGRO MALUENDA

Florianópolis, Agosto de 2005.

*Este trabalho é dedicado ao meu filho
Matías.*

Agradecimentos

Ao professor José Carlos Moreira Bermudez, pela orientação, a compreensão e a confiança durante o decorrer deste trabalho.

A meu marido Diego, pela paciência, o apoio, a força e a compreensão que teve durante este tempo todo.

À Universidade de Antofagasta–Chile e ao Projeto Mecesus Ant.01, pela oportunidade, incentivo e financiamento.

A meus colegas de laboratório, especialmente Eduardo e Guilherme pela sua boa disposição para comigo, pelo carinho e amizade.

Finalmente, gostaria também de agradecer a minha mãe e minha tia, pelo carinho e a confiança que sempre tiveram em mim.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

PROPRIEDADES DO ALGORITMO LMS OPERANDO EM PRECISÃO FINITA

Yasmín Romina Montenegro Maluenda

Agosto/2005

Orientador: José Carlos Moreira Bermudez, Ph.D.

Área de concentração: comunicações e processamento de sinais.

Palavras-chave: Filtragem adaptativa, algoritmo LMS, aritmética de precisão finita, análise estocástica, cadeia de Markov.

Número de páginas: 153.

RESUMO: Este trabalho apresenta uma nova modelagem do comportamento do algoritmo LMS quando implementado em precisão finita. Os coeficientes do filtro adaptativo são modelados como uma cadeia de Markov, e a matriz de probabilidades de transição da cadeia é determinada para o caso unidimensional. Desta forma são eliminadas as linearizações empregadas nos modelos existentes na literatura. A partir desse resultado, são determinadas as condições sobre o passo de adaptação para que o algoritmo convirja para o ponto ótimo em regime permanente. Comportamentos não observados no caso de precisão infinita podem ser claramente identificados com o modelo proposto. Resultados de simulação Monte Carlo comprovaram a qualidade do modelo proposto.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

PROPERTIES OF THE LMS ALGORITHM WORKING IN FINITE PRECISION

Yasmín Romina Montenegro Maluenda

Julho/2005

Advisor: José Carlos Moreira Bermudez, Ph.D.

Area of concentration: communications and signal processing.

Keywords: Adaptive filtering, LMS algorithm, finite-precision arithmetic, stochastic analysis, Markov chain.

Number of pages: 153.

This paper presents a new model for the behavior of the LMS algorithm when implemented in finite precision. The adaptive filter coefficients are modelled as a Markov chain and the matrix of transition probabilities is determined for the one-dimensional case. The linearization used in other models available in the literature are eliminated. Using these results, conditions on the step size for convergence to the optimum steady-state solution are determined. Algorithm behaviors that do not happen in the infinite precision case can be clearly identified with the proposed model. Monte Carlo simulation results illustrate the quality of the proposed model.

Sumário

Lista de Figuras

ix

1	INTRODUÇÃO: FILTRAGEM ADAPTATIVA E OS EFEITOS DA PRECISÃO FINITA	1
1.1	INTRODUÇÃO	1
1.2	FILTRAGEM ADAPTATIVA	1
1.2.1	Introdução	1
1.2.2	Características dos filtros adaptativos	2
1.2.3	Aplicações	4
1.2.4	Filtro de Wiener	8
1.2.5	Método do gradiente, algoritmo <i>steepest descent</i>	11
1.2.6	Algoritmo LMS (Least Mean Square)	13
1.3	EFEITOS DA PRECISÃO FINITA	15
1.3.1	Introdução	15
1.3.2	Implementação em Processadores Digitais de Sinais, DSP	16
1.3.3	Erros de Quantização	18
1.3.4	Revisão bibliográfica	19
1.4	METODOLOGIA UTILIZADA	21
1.5	DIVISÃO DE CAPÍTULOS	22
1.6	OBJETIVOS DO TRABALHO	22
1.6.1	Objetivo geral	22
1.6.2	Objetivos específicos	23
1.7	CONTRIBUIÇÕES DO TRABALHO	23
2	REVISÃO DE VARIÁVEL ALEATÓRIA E CADEIAS DE MARKOV	25
2.1	INTRODUÇÃO	25
2.2	VARIÁVEL ALEATÓRIA	26
2.2.1	Definição de variável aleatória	26
2.2.2	Função Distribuição e Função Densidade	28
2.2.3	A variável aleatória normal	31
2.2.4	Outras distribuições e densidades	35
2.2.5	Distribuição e densidade da soma de variáveis aleatórias	36
2.2.6	Funções de distribuição e densidade condicionadas	38
2.2.7	Valor esperado de uma variável aleatória	39
2.3	CADEIAS DE MARKOV	40
2.3.1	Definição e propriedades básicas	40

2.3.2	Matriz de probabilidade de transição	42
2.3.3	Equação de Chapman-Kolmogorov	43
2.3.4	Probabilidades estacionárias e tipos de cadeias de Markov	44
2.3.5	Determinação das probabilidades estacionárias	55
2.4	CONCLUSÕES	59
3	ALGORITMO LMS EM PRECISÃO FINITA	60
3.1	INTRODUÇÃO	60
3.2	CONSIDERAÇÕES INICIAIS	61
3.2.1	Equações do algoritmo LMS em precisão finita	63
3.3	DETERMINAÇÃO DA ESTATÍSTICA DE $w(n)$	63
3.3.1	Determinação da fdp de $d_Q(n)$	64
3.3.2	Determinação da fdp condicional de $e_Q(n)$	66
3.3.3	Determinação da fdp de $y_Q(n)$	68
3.3.4	Determinação da fdp de $w(n+1)$	70
3.4	RESUMO: MODELO PROPOSTO PARA $f_w\{w(n+1) x(n)=k\Delta, w(n)\}$	71
3.5	VERIFICAÇÃO DO MODELO PROPOSTO	74
3.5.1	Exemplos para $f_d(d)$	74
3.5.2	Exemplos para $f_{e_Q}(e_Q)$	77
3.5.3	Exemplos para $f_w(w)$	79
3.5.4	Exemplos para Matriz de probabilidade para $w(n)$, P_w	82
3.5.5	Probabilidade dos estados de $w(n)$ no tempo	86
3.6	CONCLUSÕES	92
4	LIMITES DO PASSO DE ADAPTAÇÃO	93
4.1	INTRODUÇÃO	93
4.2	O DIAGRAMA DE TRANSIÇÃO DOS ESTADOS DE $w(n)$	94
4.3	PASSO DE ADAPTAÇÃO- GARANTIA DE ESTADOS ABSORVENTES	98
4.4	LIMITE INFERIOR DO PASSO DE ADAPTAÇÃO	99
4.5	LIMITE SUPERIOR DO PASSO DE ADAPTAÇÃO	100
4.6	Exemplos	103
4.6.1	Caso 1:	103
4.6.2	Caso 2:	108
4.7	PASSO DE ADAPTAÇÃO E POTÊNCIA DE RUÍDO	116
4.7.1	Limite inferior do passo de adaptação em função de σ	116
4.7.2	Limite superior do passo de adaptação em função de σ	118
4.7.3	Exemplos	120
4.7.4	Passo de adaptação para valores práticos de bits ($b \geq 12$)	124
4.8	CONCLUSÕES	126
5	CONCLUSÕES	127
5.1	PROPOSTAS PARA TRABALHOS FUTUROS	128
A	Função erf	130

B	Exemplos para \hat{w} diferentes dos estados de $w(n)$	131
B.1	Caso 1	131
B.2	Caso 2	132
B.3	Caso 3	133
B.4	Caso 4	134
B.5	Caso 5	135
B.6	Caso 6	136
C	Discussão do passo de adaptação para $\hat{w} \leq -1 + \Delta$	138
C.1	Caso para $\hat{w} = -1$	138
C.1.1	Limite inferior do passo de adaptação	138
C.1.2	Limite superior do passo de adaptação	139
C.2	Caso para $\hat{w} = -1 + \Delta$	140
C.2.1	Limite inferior do passo de adaptação	140
C.2.2	Limite superior do passo de adaptação	141
C.3	Exemplos	142
C.3.1	Caso 1	142
C.3.2	Caso 2	143
C.3.3	Caso 3	143
D	Exemplos para o caso multidimensional	144
D.1	Exemplo 1	144
D.1.1	Caso 1:	145
D.1.2	Caso 2:	145
D.1.3	Caso 3:	146
D.1.4	Caso 4:	146
D.2	Exemplo 2	147
D.2.1	Caso 1:	147
D.2.2	Caso 2:	148
D.2.3	Caso 3:	148
D.2.4	Caso 4:	149
E	Símbolos e Abreviaturas	150
	Referências Bibliográficas	149

Lista de Figuras

1.1	Diagrama em blocos do problema de filtragem adaptativo	4
1.2	Diagrama em blocos para a identificação de sistemas	5
1.3	Diagrama em blocos para modelagem inversa	6
1.4	Diagrama em blocos para predição linear	6
1.5	Diagrama em blocos para cancelamento de interferência	7
1.6	Diagrama em blocos para o problema de filtragem adaptativa	8
1.7	Representação numérica nos DPS comerciais	17
1.8	(Acima) Formato de ponto flutuante IEEE-754 (abaixo) Formato em ponto fixo	17
1.9	Relação entrada-saída para um quantizador uniforme	18
2.1	Ilustração gráfica do mapeamento de uma variável aleatória [2]	26
2.2	Ilustração gráfica do mapeamento de uma variável aleatória para o Exemplo 2.1.2 [13]	27
2.3	Função distribuição (a) e função densidade (b) (baseado em [13])	30
2.4	Função densidade (a) e função distribuição (b) para uma variável aleatória gaussiana.	33
2.5	Região no plano xy onde $x + y \leq w$ [13]	37
2.6	Diagrama de Transição de estados da matriz P_y (tomado de [15])	49
2.7	Propriedades de classes em uma cadeia de Markov (tomado de [15])	51
2.8	Periodicidade das classes finais a), b) e c) (tomado de [15])	52
2.9	Exemplos de classificação de cadeias de Markov (tomado de [15])	53
3.1	Identificação de sistemas implementada em precisão finita	61
3.2	Relação entrada e saída para um quantizador de 3 bits	62
3.3	Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 2bits$, $\sigma_z =$ 10^{-1} e $\hat{w} = 0,25$	75
3.4	Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 2bits$, $\sigma_z =$ 10^{-3} e $\hat{w} = 0,25$	75
3.5	Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 3bits$, $\sigma_z =$ 10^{-3} e $\hat{w} = 0,75$	76
3.6	Comparação entre o Modelo e a Simulação para $f_{eQ}(e_Q)$ com $b = 4bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 1,425$ e $w(0) = 0,125$	77
3.7	Comparação entre o Modelo e a Simulação para $f_{eQ}(e_Q)$ com $b = 4bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 1$ e $w(0) = -1$	78
3.8	Comparação entre o Modelo e a Simulação para $f_{eQ}(e_Q)$ com $b = 5bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 1,425$ e $w(0) = -0,03125$	78

3.9	Comparação entre o Modelo e a Simulação para $f_{eQ}(e_Q)$ com $b = 4bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 0$ e $w(0) = -1$	79
3.10	Simulação versus equações (3.21) e (3.22), para $b = 4$ bits, $\mu = 0,5$, $\hat{w} = 0,875$, $w(0) = -1$ e $\sigma = 10^{-3}$	80
3.11	Simulação versus equações (3.21) e (3.22), $b = 5$ bits, $\mu = 0,7$, $\hat{w} = 0,5625$, $w(0) = -0,125$ e $\sigma = 10^{-3}$	80
3.12	Simulação versus equações (3.21) e (3.22), para $b = 6$ bits, $\mu = 0,8$, $\hat{w} = 0,40625$, $w(0) = -0,78125$ e $\sigma = 10^{-3}$	81
3.13	Simulação versus equações (3.21) e (3.22), para $b = 7$ bits, $\mu = 1$, $\hat{w} = 0,703125$, $w(0) = -0,15625$ e $\sigma = 10^{-5}$	81
3.14	Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 1$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = -0,25$	87
3.15	Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 0,58$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = 0,25$	89
3.16	Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 0,2$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = 0,75$	91
4.1	Diagrama de transição de estados de $w(n)$ com $b = 3bits$, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 1$	95
4.2	Diagrama de transição de estados de $w(n)$ com $b = 3bits$, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,58$	96
4.3	Diagrama de transição de estados de $w(n)$ com $b = 3$ bits, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,2$	97
4.4	Ilustração do método de quantização	117
D.1	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0; 0; 0; 0]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$	145
D.2	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-0,50; -0,25; 0,00; 0,25]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$	145
D.3	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-1; -1; -1; -1]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$	146
D.4	Comportamento algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0,25; 0,5; -0,25; 0]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$	146
D.5	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0; 0; 0; 0]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$	147
D.6	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-0,50; -0,25; 0,00; 0,25]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$	148
D.7	Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-1; -1; -1; -1]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$	148
D.8	Comportamento algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0,25; 0,5; -0,25; 0,00]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$	149

Capítulo 1

INTRODUÇÃO: FILTRAGEM ADAPTATIVA E OS EFEITOS DA PRECISÃO FINITA

1.1 INTRODUÇÃO

Neste capítulo são introduzidos os conceitos iniciais para melhor entendimento do trabalho e apresentados os objetivos geral e específicos. A seguir apresenta-se uma breve revisão sobre os princípios da filtragem adaptativa, dando ênfase ao algoritmo adaptativo LMS e aos efeitos da precisão finita na sua implementação.

1.2 FILTRAGEM ADAPTATIVA

1.2.1 Introdução

A filtragem adaptativa constitui atualmente uma ferramenta importante no processamento estatístico de sinais, especialmente quando precisa-se processar sinais provenientes de ambientes com estatísticas desconhecidas ou que variam no tempo. O estudo da filtragem adaptativa foi impulsionado com o desenvolvimento do algoritmo Least Mean Square (LMS) por Widrow e Hopf em 1960 e tem recebido uma especial atenção por muitos pesquisadores nas últimas décadas, devido ao interesse de resolver problemas práticos que não podem ser resolvidos satisfatoriamente utilizando filtros fixos.

Os filtros adaptativos são projetados com base em algoritmos recursivos, os quais ajustam seus parâmetros com o intuito de minimizar uma função custo. O filtro é inicializado com um conjunto qualquer de coeficientes, dado o desconhecimento do ambiente de operação. Após um determinado número de iterações, o algoritmo adaptativo converge levando os coeficientes do filtro para valores suficientemente próximos da solução ótima. Em ambientes não estacionários o algoritmo permite ao filtro a possibilidade de acompanhar as variações estatísticas dos sinais desde que sejam suficientemente lentas, qualidade conhecida como rastreamento.

Os rápidos avanços no desenvolvimento de processadores de sinal mais velozes, tornou possível a aplicação e a implementação de sistemas adaptativos em áreas tão diversas como comunicações, radar, sismologia, mecânica, acústica e engenharia biomédica [1]. Apesar de serem aplicações de naturezas tão diferentes, todas possuem uma característica comum: requerem a estimação em tempo real de um sinal desejado a partir da observação de sinais de referência e do erro de estimação. Esses sinais são empregados pelo sistema adaptativo para controlar os coeficientes do filtro.

1.2.2 Características dos filtros adaptativos

Existe uma grande variedade de algoritmos recursivos desenvolvidos para adaptar os coeficientes dos filtros adaptativos, a escolha de um ou outro algoritmo é determinada por um ou mais dos seguintes fatores:

- *Taxa de convergência*: Definida como o número de iterações necessárias para que o algoritmo consiga levar os coeficientes do filtro a valores suficientemente próximos da solução ótima.
- *Desajuste*: Medida da diferença entre os parâmetros obtidos após um certo número de iterações e os parâmetros ótimos.

- *Rastreamento*: Capacidade do algoritmo de acompanhar as variações das propriedades estatísticas de sinais não estacionários.
- *Robustez*: Habilidade do algoritmo para operar satisfatoriamente com um dado de entrada mal condicionado.
- *Complexidade computacional* [1]: Refere-se a: 1) número de operações aritméticas requeridas para fazer uma iteração completa do algoritmo; 2) tamanho da memória requerida e 3) flexibilidade da programação.
- *Estrutura* [1]: Refere-se à estrutura do fluxo da informação no algoritmo e a maneira de ser implementada no hardware.
- *Propriedades numéricas*: A implementação do algoritmo adaptativo em um processador digital pode fazer com que o algoritmo se desvie da operação ideal, determinada em precisão infinita, produzindo acumulação de erros de quantização. É possível que os erros se acumulem sem limite. Se esta situação persistir sem controle, diz-se que o algoritmo é numericamente instável.

O filtro adaptativo é um filtro cujos coeficientes são atualizados por um algoritmo adaptativo para otimizar a resposta do filtro segundo um critério de desempenho. Em geral os filtros adaptativos consistem de duas partes: 1) o processo de filtragem, cuja saída é a resposta desejada para um ou vários sinais de entrada; 2) o processo adaptativo, que provê um mecanismo de controle adaptativo, através de algum algoritmo, para o ajuste dos parâmetros usados no processo de filtragem.

O problema de filtragem adaptativa é apresentado de forma esquemática na Figura 1.1

Nesta figura:

$d(n)$: é a resposta ou o sinal desejado

$x(n)$: é o sinal de entrada (referência)

$y(n)$: é a saída do filtro adaptativo, que é uma estimativa de $d(n)$

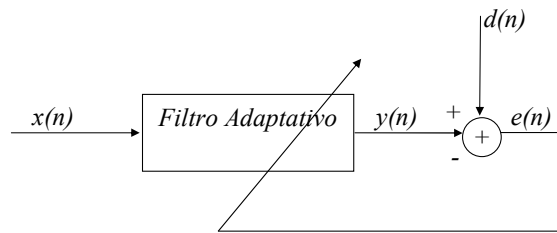


Figura 1.1: Diagrama em blocos do problema de filtragem adaptativo

$e(n) = d(n) - y(n)$: é o sinal de erro

Os parâmetros do filtro são atualizados de n para $n + 1$ pelo algoritmo adaptativo de forma a minimizar uma figura de mérito baseada no erro de estimação $e(n)$.

Os filtros adaptativos podem ser implementados usando estruturas com respostas ao impulso de duração finita ou infinita. Assim, têm-se os filtros FIR (Finite Impulse Response) e os filtros IIR (Infinite Impulse Response), respectivamente. Neste trabalho a discussão será restrita aos filtros adaptativos FIR porque, ao contrário dos filtros IIR, são sempre estáveis, já que todos os pólos da sua função de transferência estão localizados na origem do plano z . Além disso, filtros FIR podem fornecer resposta com fase linear. Os filtros IIR podem tornar-se instáveis, o que complica sobremaneira o controle do algoritmo adaptativo e o tratamento matemático de análise de seu desempenho.

1.2.3 Aplicações

Os filtros adaptativos têm sido aplicados com muito sucesso em diversas áreas. Todas estas aplicações têm em comum um vetor de entrada e um sinal desejado, os quais são utilizados juntamente com o sinal de erro para ajustar os coeficientes do filtro de forma a otimizar o seu desempenho. A diferença essencial entre as várias aplicações é a forma de extrair a resposta desejada, assim as aplicações dos filtros adaptativos podem classificar-se em:

- *Identificação de sistemas*: Nesta classe de aplicação, o filtro adaptativo é usado para

prover o modelo linear que melhor represente uma planta desconhecida. A Figura 1.2 apresenta um diagrama em blocos para este tipo de aplicação. A planta e o filtro adaptativo são alimentados pelo mesmo sinal de entrada e a saída da planta é a resposta desejada para o filtro adaptativo. O ruído aditivo $z(n)$, representa o ruído interno da planta ou ruído de medição. Fazendo a adaptação dos coeficientes do filtro através de um algoritmo recursivo obtém-se uma boa aproximação para o modelo da planta desconhecida. A identificação adaptativa é freqüentemente utilizada em sistemas de comunicações, sistemas de controle e em identificação de estruturas. Por exemplo, nos sistemas de controle pode ser utilizada para modelar uma planta com a finalidade de projetar um controlador.

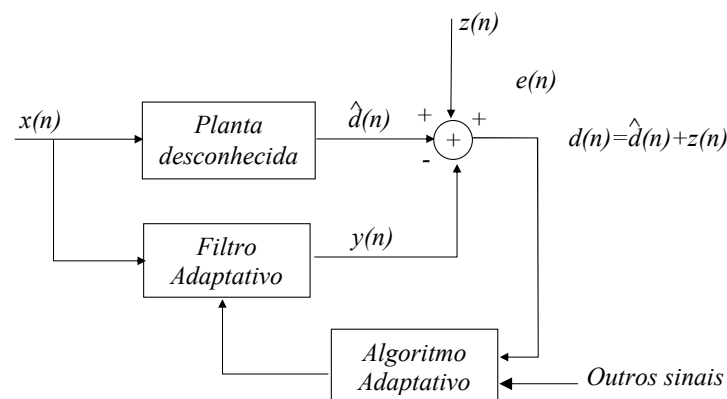


Figura 1.2: Diagrama em blocos para a identificação de sistemas

- *Modelagem inversa:* A função do filtro adaptativo nesta aplicação é prover um modelo inverso de um sistema desconhecido. Idealmente, a função de transferência do modelo inverso representa a função de transferência inversa do sistema desconhecido, a combinação de ambas funções constitui um meio de transmissão ideal. O sinal desejado para o filtro adaptativo é a entrada atrasada do sistema. Uma aplicação da modelagem inversa é a equalização de canal, em que o objetivo é reduzir a interferência intersimbólica através de um filtro adaptativo que inverta a resposta do canal. Observe na Figura 1.3 um diagrama em blocos da modelagem inversa.

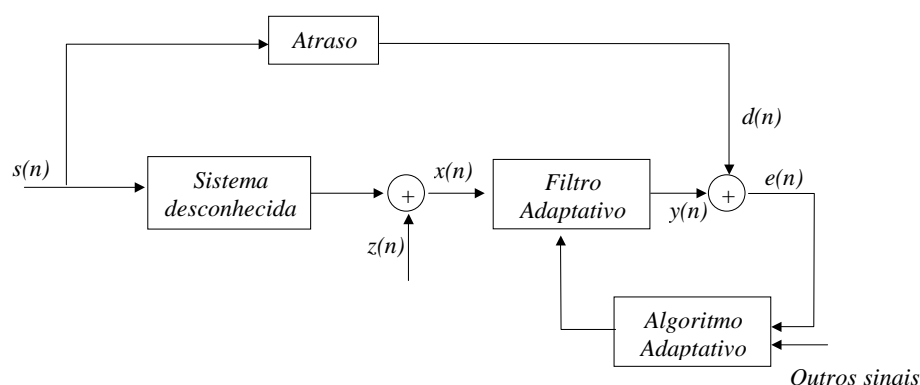


Figura 1.3: Diagrama em blocos para modelagem inversa

- *Predição linear*: Nesta aplicação a função do filtro adaptativo é prover a melhor predição do valor presente de um sinal aleatório. O valor presente serve como sinal desejado no filtro adaptativo e os valores passados alimentam o filtro. Dependendo da aplicação, pode-se ter duas saídas: $y(n)$, denominada saída do filtro preditor ou $e(n)$, denominada saída do filtro de erro de predição. Uma de suas aplicações é na análise espectral, em que utiliza-se o modelamento preditivo para estimar a potência espectral de um sinal de interesse [1]. A Figura 1.4 apresenta o diagrama em blocos para um preditor linear.

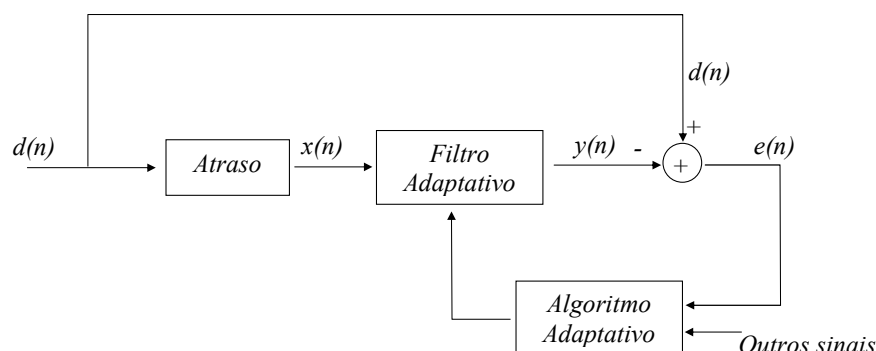


Figura 1.4: Diagrama em blocos para predição linear

- *Cancelamento de interferência*: O filtro adaptativo é utilizado para cancelar uma interferência desconhecida contida em um sinal primário. Este sinal primário serve como

sinal desejado para o filtro adaptativo. Um sinal correlacionado com a interferência alimenta a entrada do filtro. Utiliza-se em áreas tão diversas como o cancelamento de interferência de 50Hz ou 60Hz da rede elétrica em electrocardiogramas, cancelamento de ruído em sinais de fala, etc. Observe na Figura 1.5 o diagrama em blocos para esta aplicação.

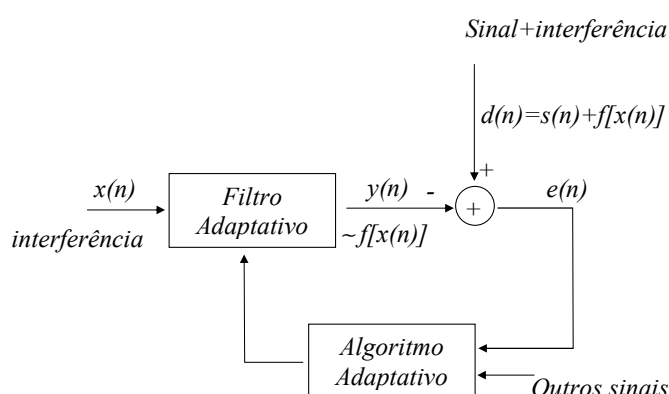


Figura 1.5: Diagrama em blocos para cancelamento de interferência

O problema de filtragem adaptativa não tem uma solução única, já que existe uma grande variedade de algoritmos recursivos. De uma maneira geral, a maioria dos algoritmos adaptativos pode ser subdividida em duas classes:

- *Algoritmos baseados no método de gradiente*: Estes algoritmos baseiam-se no filtro de Wiener, estimando o gradiente da superfície de uma função custo, baseada no erro médio quadrático (valor médio quadrático da diferença entre a resposta desejada e a saída do filtro). No caso de implementações usando estruturas de filtragem FIR transversais, a dependência do erro médio quadrático com os coeficientes assume a forma de uma superfície parabolóide com um único ponto mínimo. Os coeficientes que determinam o ponto mínimo da superfície definem a solução ótima de Wiener. Algoritmos dessa classe têm como principais qualidades a baixa complexidade computacional e robustez a variações nas condições de operação.
- *Algoritmos baseado no método dos mínimos quadrados*: Os algoritmos de mínimos

quadrados são baseados no filtro de Kalman. De forma determinística minimizam a soma dos quadrados dos erros parciais dentro de uma janela de observação. Têm como vantagem a baixa sensibilidade a mínimos locais das superfícies da função custo e uma maior velocidade de convergência (medida em número de iterações) comparados com os algoritmos do gradiente. Suas principais desvantagens são a complexidade computacional elevada e problemas de robustez, estabilidade e propriedades numéricas.

Considerando que o interesse deste trabalho é o algoritmo LMS (Least Mean Square), e que este é um algoritmo baseado no método do gradiente, nas seções seguintes apresenta-se uma breve revisão do filtro de Wiener, do método do gradiente e do algoritmo LMS.

1.2.4 Filtro de Wiener

Considere a Figura 1.6, em que a entrada ao filtro consiste da sequência $\{x(0), x(1), x(2), \dots\}$ e o filtro caracteriza-se pela resposta ao impulso $\{w_0, w_1, w_2, \dots\}$. No tempo discreto n , o filtro produz a saída $y(n)$. Utiliza-se esta saída como uma estimativa da resposta desejada $d(n)$. O erro de estimação $e(n)$ é dado pela diferença entre a resposta desejada $d(n)$ e a saída do filtro $y(n)$. O objetivo é determinar o conjunto de coeficientes w_k do filtro linear, de forma a minimizar o valor médio quadrático do erro de estimação.

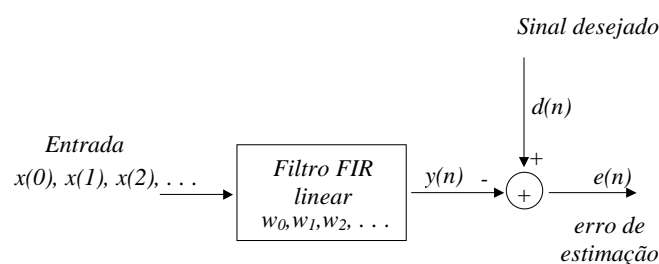


Figura 1.6: Diagrama em blocos para o problema de filtragem adaptativa

Então com,

$$\begin{aligned} e(n) &= d(n) - y(n) \\ y(n) &= \sum_{k=0}^{N-1} w_k x(n-k) \\ J_{ms}(n) &= E\{e^2(n)\} \quad \text{valor médio quadrático do erro de estimação} \end{aligned}$$

e definindo:

$$\begin{aligned} \underline{\mathbf{X}}(n) &= [x(n), x(n-1), x(n-2), \dots, x(n-N+1)] \quad \text{vetor de entrada} \\ \underline{\mathbf{W}} &= [w_0, w_1, w_2, \dots, w_{N-1}] \quad \text{vetor dos coeficientes} \end{aligned}$$

tem-se que,

$$\begin{aligned} y(n) &= \underline{\mathbf{X}}^T(n) \underline{\mathbf{W}} \\ e(n) &= d(n) - \underline{\mathbf{X}}^T(n) \underline{\mathbf{W}} \\ e^2(n) &= d^2(n) - 2d(n)\underline{\mathbf{X}}^T(n)\underline{\mathbf{W}} + \underline{\mathbf{W}}^T \underline{\mathbf{X}}(n) \underline{\mathbf{X}}^T(n) \underline{\mathbf{W}} \end{aligned}$$

Assim, o erro médio quadrático é dado por,

$$\begin{aligned} J_{ms}(n) &= E[e^2(n)] \\ &= E[d^2(n)] - 2E[d(n)\underline{\mathbf{X}}^T(n)]\underline{\mathbf{W}} + \underline{\mathbf{W}}^T E[\underline{\mathbf{X}}(n)\underline{\mathbf{X}}^T(n)]\underline{\mathbf{W}} \end{aligned}$$

Nesta expressão, a seguinte notação é comumente empregada:

$$\begin{aligned} \underline{\mathbf{P}}_{xd}(n) &= E[d(n)\underline{\mathbf{X}}(n)] \quad \text{vetor de correlação cruzada entre } d(n) \text{ e } \underline{\mathbf{X}}(n) \\ \underline{\mathbf{R}}_{xx}(n) &= E[\underline{\mathbf{X}}(n)\underline{\mathbf{X}}^T(n)] \quad \text{matriz de autocorrelação de } \underline{\mathbf{X}}(n) \end{aligned}$$

Quando $d(n)$ e $\underline{\mathbf{X}}(n)$ são processos estacionários, $\underline{\mathbf{P}}_{xd}$ e $\underline{\mathbf{R}}_{xx}$ são invariantes em n . Assumindo que $x(n)$ tem média zero e escrevendo $E[d^2(n)] = \sigma_d^2$, que é a variância da resposta desejada,

terá-se:

$$J_{ms} = \sigma_d^2 - 2\mathbf{P}_{xd}^T \mathbf{W} + \mathbf{W}^T \mathbf{R}_{xx} \mathbf{W} \quad (1.1)$$

A equação (1.1), é uma função quadrática em \mathbf{W} com forma de hiperparabolóide de concavidade para cima e um único mínimo, também conhecida como *Superfície de Desempenho*.

O gradiente da função (1.1) é,

$$\nabla J_{ms} = \frac{\partial J_{ms}}{\partial \mathbf{W}}$$

em que,

$$\begin{aligned} \frac{\partial \mathbf{P}_{xd}^T \mathbf{W}}{\partial \mathbf{W}} &= \mathbf{P}_{xd} \\ \frac{\partial (\mathbf{W}^T \mathbf{R}_{xx} \mathbf{W})}{\partial \mathbf{W}} &= \mathbf{R}_{xx} \mathbf{W} + \mathbf{R}_{xx}^T \mathbf{W} = 2\mathbf{R}_{xx} \mathbf{W} \end{aligned}$$

Assim,

$$\nabla J_{ms} = -2\mathbf{P}_{xd} + 2\mathbf{R}_{xx} \mathbf{W} \quad (1.2)$$

O vetor de pesos ótimo $\hat{\mathbf{W}}$ é aquele para o qual o gradiente anula-se, ou seja

$$\nabla J_{ms} = -2\mathbf{P}_{xd} + 2\mathbf{R}_{xx} \hat{\mathbf{W}} = \mathbf{0} \quad (1.3)$$

Assumindo que \mathbf{R}_{xx}^{-1} existe, tem-se que o vetor ótimo dos coeficientes ou a *solução de Wiener* no sentido médio quadrático é,

$$\hat{\mathbf{W}} = \mathbf{R}_{xx}^{-1} \mathbf{P}_{xd} \quad (1.4)$$

Substituindo o resultado anterior na equação (1.1) tem-se que o mínimo erro médio quadrático (MSE, Mean-Square Error) é dado por:

$$J_{min} = \sigma_d^2 - \hat{\mathbf{W}}^T \mathbf{R}_{xx} \hat{\mathbf{W}} \quad (1.5)$$

Modelando o sinal desejado $d(n)$ como $d(n) = \mathbf{X}^T(n) \hat{\mathbf{W}} + \eta(n)$ e sendo $\eta(n)$ um ruído branco de média zero e estatisticamente independente de $\mathbf{X}(n)$, tem-se:

$$J_{min} = \sigma_\eta^2$$

em que $\sigma_\eta^2 = E[\eta^2(n)]$. O projeto do filtro de Wiener dado pela equação (1.4) requer o conhecimento *a priori* das estatísticas dos dados a processar ($\mathbf{R}_{xx}, \mathbf{P}_{xd}$). Em geral, essas estatísticas não são conhecidas, devendo ser estimadas a partir dos dados disponíveis.

1.2.5 Método do gradiente, algoritmo *steepest descent*

O método do gradiente é uma técnica recursiva para a determinação do ponto mínimo de uma função custo. A determinação da solução ótima é feita por um sistema realimentado que aprimora a cada passo a estimativa dos coeficientes ótimos [1]. Inicia-se com um valor arbitrário do vetor de coeficientes \mathbf{W} , o qual é ajustado recursivamente. A solução converge para a solução de Wiener sem a necessidade de inverter a matriz de autocorrelação do vetor de entrada.

Para encontrar uma solução recursiva, o vetor de coeficientes passa a ser variante no tempo. Assim,

$$\mathbf{W} \Rightarrow \mathbf{W}(n) \quad (1.6)$$

$$J_{ms} \Rightarrow J_{ms}(n) \quad (1.7)$$

Substituindo as equações (1.6) e (1.7) em (1.1) tem-se que,

$$J_{ms}(n) = \sigma_d^2 - 2\mathbf{P}_{xd}^T \mathbf{W}(n) + \mathbf{W}(n)^T \mathbf{R}_{xx} \mathbf{W}(n) \quad (1.8)$$

O gradiente da superfície $J_{ms}(n)$ na iteração n é dado por

$$\nabla J_{ms}(n) = -2\mathbf{P}_{xd} + 2\mathbf{R}_{xx} \mathbf{W}(n) \quad (1.9)$$

Este é o vetor que aponta na direção de deslocamento no espaço \mathbf{W} para que $J_{ms}(\mathbf{W})$ tenha o maior aumento possível. Considere para a atualização do vetor \mathbf{W} a expressão

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \frac{1}{2}\mu\mathbf{P}(n) \quad (1.10)$$

em que $\mathbf{P}(n)$ é o vetor que estabelece a direção do ajuste em $\mathbf{W}(n)$ e μ é o passo de atualização. Para a escolha de $\mathbf{P}(n)$ considera-se que $J_{ms}(n+1) < J_{ms}(n)$ é a condição necessária para alcançar a convergência quando $n \rightarrow \infty$. Além disso, sabendo que

$$J_{ms}(n+1) = \sigma_d^2 - 2\mathbf{P}_{xd}^T \mathbf{W}(n+1) + \mathbf{W}(n+1)^T \mathbf{R}_{xx} \mathbf{W}(n+1)$$

e substituindo a atualização de $\mathbf{W}(n+1)$, de [2] tem-se que:

$$\mathbf{P}(n) = -\nabla J_{ms}(n)$$

substituindo $\mathbf{P}(n)$ na equação (1.10) a equação de atualização de $\mathbf{W}(n)$ para o método do gradiente fica

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) - \frac{1}{2}\mu\nabla J_{ms}(n) \\ \mathbf{W}(n+1) &= \mathbf{W}(n) + \mu[\mathbf{P}_{xd} + \mathbf{R}_{xx} \mathbf{W}(n)] \end{aligned} \quad (1.11)$$

Note que, embora as estimativas de \mathbf{P}_{xd} e \mathbf{R}_{xx} ainda sejam necessárias, o processo recursivo

elimina a necessidade de inversão de \underline{R}_{xx} .

1.2.6 Algoritmo LMS (Least Mean Square)

O algoritmo LMS, denominado assim por Widrow e Hoff em 1960, é um dos mais importantes da família dos algoritmos de gradiente estocástico [1]. A característica mais importante deste algoritmo é a simplicidade, representada por uma fácil implementação e uma baixa complexidade computacional. Além disso, não requer o conhecimento das estatísticas dos sinais envolvidos, sendo capaz de rastrear mudanças no ambiente de operação.

Diferentemente do método do gradiente, o algoritmo LMS não requer o conhecimento das estatísticas \underline{P}_{xd} e \underline{R}_{xx} . Sua equação de atualização dos coeficientes baseia-se em estimativas instantâneas dessas estatísticas. Isso facilita a implementação, mas transforma a equação de atualização dos coeficientes em uma equação estocástica. Os coeficientes do filtro adaptativo passam a ser variáveis aleatórias.

Estimando \underline{P}_{xd} e \underline{R}_{xx} a partir de valores instantâneos \underline{P}_{xd} e \underline{R}_{xx} temos

$$\hat{\underline{R}}_{xx} = \underline{X}(n)\underline{X}^T(n)$$

$$\hat{\underline{d}}_{xd} = d(n)\underline{X}(n)$$

considerando a equação (1.9), a estimativa instantânea do gradiente fica,

$$\hat{\nabla}\underline{J}_{ms}(n) = -2d(n)\underline{X}(n) + 2\underline{X}(n)\underline{X}^T(n)\underline{W}(n) \quad (1.12)$$

então,

$$\underline{W}(n+1) = \underline{W}(n) - \frac{1}{2}\mu\hat{\nabla}\underline{J}_{ms}(n) \quad (1.13)$$

Substituindo a equação (1.12) em (1.13) tem-se que,

$$\begin{aligned}\underline{\mathbf{W}}(n+1) &= \underline{\mathbf{W}}(n) - \frac{1}{2}\mu[-2d(n)\underline{\mathbf{X}}(n) + 2\underline{\mathbf{X}}(n)\underline{\mathbf{X}}^T(n)\underline{\mathbf{W}}(n)] \\ &= \underline{\mathbf{W}}(n) - \mu[-d(n) + \underline{\mathbf{X}}^T(n)\underline{\mathbf{W}}(n)]\underline{\mathbf{X}}(n) \\ &= \underline{\mathbf{W}}(n) + \mu e(n)\underline{\mathbf{X}}(n)\end{aligned}\tag{1.14}$$

em que $-e(n)\underline{\mathbf{X}}(n)$ é a estimativa ruidosa do gradiente e μ é o passo de atualização que controla a estabilidade e a convergência do algoritmo.

Demonstra-se em [2] que, uma condição suficiente para estabilidade do algoritmo LMS é:

$$0 < \mu < \frac{2}{N\sigma_x^2}\tag{1.15}$$

em que N é a ordem do filtro e σ_x^2 a potência de $x(n)$.

Em [1] e [2] mostra-se que a evolução da função custo $J_{ms}(n)$ do algoritmo LMS pode ser aproximada por:

$$J_{ms}(n) = J_{min} + tr\{\underline{\mathbf{R}}_{xx}\underline{\mathbf{K}}(n)\}$$

em que $tr\{\cdot\}$ corresponde ao traço da matriz e considerando as hipótese feitas sobre $d(n)$ e $\underline{\mathbf{X}}(n)$ após (1.3).

Os passos do algoritmo LMS são resumidos na Tabela 1.1.

O algoritmo LMS é um dos algoritmos mais populares utilizados na implementação digital de filtros adaptativos [3]. Por isso pesquisas como as feitas em [4] - [9] estudam e propõem modelos que possam prever o comportamento do algoritmo LMS quando implementado em precisão finita. Na próxima seção descreve-se com mais detalhe os efeitos da precisão finita na implementação digital do algoritmo LMS.

Tabela 1.1: Sumário do algoritmo adaptativo LMS

Dados	
$\underline{X}(n)$	= vetor de entrada no tempo n
$d(n)$	= resposta desejada no tempo n
$\underline{W}(n)$	= vetor dos coeficientes do filtro no tempo n
μ	= tamanho do passo de atualização
Inicialização	
$\underline{W}(0) = \underline{X}(0)$	$= \underline{0}$
Computação	
$y(n)$	$= \underline{X}^T(n) \underline{W}$
$e(n)$	$= d(n) - y(n)$
$\underline{W}(n+1)$	$= \underline{W}(n) + \mu e(n) \underline{X}(n)$

1.3 EFEITOS DA PRECISÃO FINITA

1.3.1 Introdução

A teoria de filtragem adaptativa frequentemente considera que todas as variáveis envolvidas nos cálculos, como as entradas e os coeficientes do filtro, podem ser representados em precisão infinita. Esta suposição facilita a análise matemática. No entanto, a implementação prática destes sistemas em um *Processador Digital de Sinais* (DSP, Digital Signal Processor), em que a aquisição e o armazenamento dos sinais realiza-se com um número finito de dígitos, pode levar a resultados que desviam-se dos cálculos teóricos devido à quantização da entrada e dos cálculos internos. A acumulação de erros de quantização pode levar o comportamento do filtro adaptativo a níveis inaceitáveis, razão pela qual precisa-se dar maior atenção aos efeitos da precisão finita.

A seguir, apresenta-se uma breve descrição de algumas das características do DSP, com a finalidade de compreender melhor o comportamento do algoritmo LMS quantizado (implementado em precisão finita), examina-se as fontes de erros de quantização e sua relação com a estabilidade e exatidão numérica e, finalmente, faz-se uma revisão bibliográfica dos estudos feitos até agora do algoritmo LMS quando implementado em precisão finita.

1.3.2 Implementação em Processadores Digitais de Sinais, DSP

Um DSP é um microprocessador especificamente projetado para realizar o processamento digital de sinais. DSPs utilizam arquiteturas especiais para acelerar os cálculos matemáticos característicos da maioria dos sistemas de processamento de sinais em tempo real.

Algumas de suas características mais básicas como formato aritmético, velocidade de processamento, organização da memória e arquitetura interna determinam sua adequação a uma aplicação em particular. São utilizados em diversas aplicações, desde sistemas de radar até eletrônica de consumo. Na atualidade, existem quatro grandes fabricantes de DSP: Texas Instruments com a série TMS320; Motorola com as séries DSP56000, DSP56100, DSP56600 e DSP96000; Lucent Technologies, com as séries DSP1600 e DSP3200; e Analog Devices, com as séries ADSP2100 e ADSP21000.

Uma das características fundamentais dos DSP é o tipo de formato aritmético utilizado pelo processador para a representação numérica. Esta representação numérica realiza-se em base 2, ou seja, na forma binária e dois formatos: formato em ponto fixo e formato em ponto flutuante. Em ponto fixo utiliza-se um número determinado de dígitos para a parte inteira e outro para a parte decimal, por exemplo: 1001,011010 (4 dígitos na parte inteira e 6 dígitos na parte decimal). Em ponto flutuante a representação é da forma: $M2^c$, em que M e c são inteiros. A Figura 1.7 mostra a relação que existe entre formato numérico e o número de bits nos DSP comerciais [10].

A Figura 1.8 apresenta um exemplo para os formatos de ponto flutuante e ponto fixo. Para o formato IEEE-754 de ponto flutuante, a letra s indica que o bit mais significativo é o sinal, a letra e indica o expoente formado por 8 bits e a letra m indica a mantissa do número (aqui com 23 bits). Como o formato de ponto fixo não tem expoente, permite representar números com mais bits significativos, para o mesmo número de bits, proporcionando uma maior resolução que o formato em ponto flutuante.

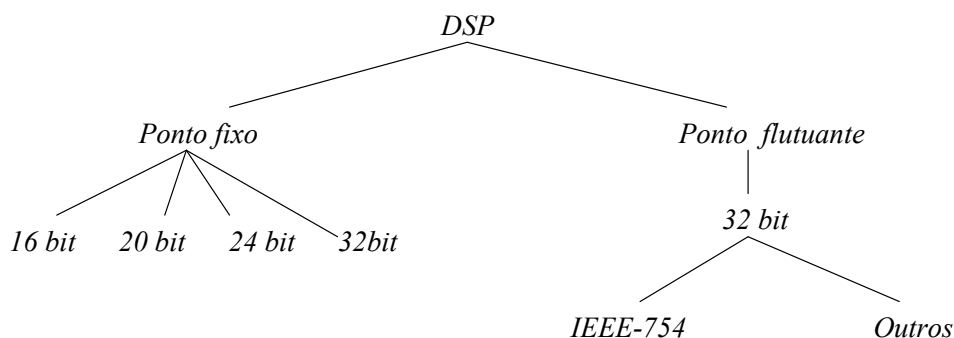


Figura 1.7: Representação numérica nos DPS comerciais

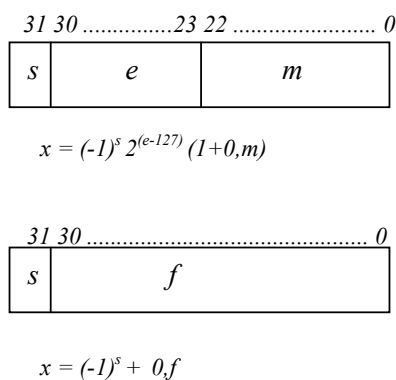


Figura 1.8: (Acima) Formato de ponto flutuante IEEE-754 (abaixo) Formato em ponto fixo

Os DSP de ponto flutuante utilizam um barramento de dados de 32 bits enquanto que nos DSP de ponto fixo o tamanho mais comum é de 16 bits. O tamanho do barramento determina o tamanho da palavra e tem um grande impacto no custo, já que influi diretamente no tamanho do "chip", assim como no tamanho da memória externa conectada aos DSP. Se a aplicação requer mais precisão, um tamanho da palavra maior será melhor, mas o custo da implementação digital do algoritmo aumenta com o número de bits. Portanto, existe uma ampla motivação para utilizar o mínimo número de bits possível [1].

Os sinais processados por um DSP são sinais representados na forma digital, ou em precisão finita. Quando os sinais não se encontram na forma digital, devem ser convertidos por conversores analógicos–digitais, provocando erros de quantização. A seguir apresentam–

se os erros que ocorrem quando implementa-se digitalmente um filtro adaptativo.

1.3.3 Erros de Quantização

Essencialmente existem duas fontes de quantização presentes na implementação digital de um filtro adaptativo:

- 1 *Conversão análogo-digital (A/D)[1]*: Um sinal deve encontrar-se em precisão finita para ser tratada no DSP. Na conversão A/D, além do processo de amostragem deve-se considerar o processo de quantização, que atribui um número limitado de níveis para a representação de cada amostra do sinal, de acordo com o tamanho dos registros do DSP que armazenarão seu valor. A representação numérica de cada amostra corresponde à saída do quantizador. A Figura 1.9 apresenta um quantizador uniforme, com tamanho do passo δ e níveis de quantização $0, \pm\delta, \pm2\delta, \dots$. Em [1] denomina-se η ao erro de quantização ou arredondamento correspondente à diferença entre a entrada e a saída do quantizador, em que $-\frac{\delta}{2} \leq \eta \leq \frac{\delta}{2}$. Observe que a relação entre a entrada e a saída do quantizador corresponde a uma função não linear.

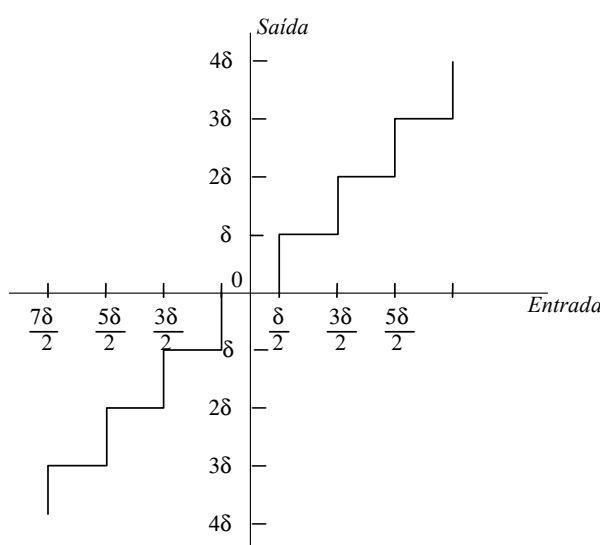


Figura 1.9: Relação entrada-saída para um quantizador uniforme

2 *Tamanho da palavra e aritmética [1]*: O tamanho da palavra é o número de bits utilizados para armazenar os resultados da aritmética interna. A operação de maior interesse é a multiplicação, já que o seu resultado ocupa o dobro do número de bits que os fatores. Assim, ao serem armazenados produz-se um erro ao reduzir à metade o número de bits. No caso da implementação digital de filtros adaptativos, os coeficientes do filtro são atualizados continuamente, o que propaga os erros de quantização e faz com que o comportamento do filtro desvie-se do ideal (teoria). Se o desvio continua aumentando, ou seja, os erros do resultado da aritmética de precisão finita acumulam-se sem limite o filtro cai em uma condição de "overflow"¹. Neste caso, diz-se que o algoritmo é *numericamente instável*. O overflow é inaceitável para qualquer aplicação na qual o filtro adaptativo opera continuamente [6]. Um filtro adaptativo é dito *numericamente estável* se os desvios dos resultados da aritmética em precisão finita, comparados com os de precisão infinita, são limitados [1].

Outro parâmetro que requer atenção na implementação digital do filtro é a *exatidão numérica*. A exatidão numérica de um filtro adaptativo é determinada pelo número de bits utilizados nos cálculos internos da implementação do filtro [1]. Quanto maior o número de bits, menor o desvio em relação ao desempenho ideal do filtro [1].

A acumulação de erros de quantização pode alcançar níveis inaceitáveis. Por isso, a implementação em precisão finita de filtros adaptativos requer especial atenção. A seguir, faz-se uma revisão bibliográfica dos estudos e modelos propostos para a análise do desempenho do algoritmo LMS quantizado.

1.3.4 Revisão bibliográfica

A Figura 1.9 mostra a relação entre entrada e saída para um quantizador. Esta operação é inerentemente não-linear [7]. Assim, o desempenho do algoritmo LMS quantizado (implementação digital), não pode ser previsto por uma análise linear em precisão infinita. Além disso, sendo

¹ Condição dos dados em operações aritméticas em que a magnitude do resultado excede o nível máximo que pode ser representado com o número de bits disponível.

a quantização uma operação não invertível², torna-se difícil obter uma análise estatística exata do comportamento do algoritmo LMS em precisão finita [12].

Vários métodos aproximados de análise têm sido propostos. Caraiscos e Liu [4] realizaram uma análise estatística de segunda ordem para aritméticas de ponto fixo e de ponto flutuante para o algoritmo LMS quantizado. No entanto, consideram um modelo linear branco não correlacionado, de média zero e variância σ_d^2 para o erro de quantização. Alexander [5] estudou o comportamento de um vetor de desajuste durante o período transiente do algoritmo LMS. Este desajuste mede a diferença entre os coeficientes obtidos com precisão infinita e os obtidos com precisão finita. Como em [4], modela-se o erro de quantização como um ruído aditivo de média zero, não correlacionado com o sinal a ser quantizado e distribuído uniformemente. Para o algoritmo LMS quantizado com coeficientes de precisão finita e entrada de precisão infinita, Bermudez e Bershad [3]-[9] apresentam uma curva de aprendizado recursiva mais exata para o EMQ (Erro médio quadrático) do que as análises anteriores. Em [3] determinam uma equação recursiva para o erro médio quadrático para um sistema estacionário. Em [7], usando um modelo de Markov de primeira ordem para um sistema variante no tempo, determinam uma equação recursiva para o EMQ no caso não-estacionário. Em [8] estuda-se o fenômeno de "slowdown"³ no LMS quantizado. Estes resultados baseiam-se em um modelo mais simplificado para o quantizador. Eles assumem um comportamento linear do quantizador a partir de entradas superiores a δ . (observe Figura 1.9). Em [9] estuda-se o comportamento do LMS quantizado tendo o passo μ como potência de dois. Isso modifica o comportamento do algoritmo LMS evitando uma multiplicação no termo de atualização dos coeficientes. A operação do quantizador é expressa como a soma de uma função linear e uma dente de serra periódica.

As análises discutidas acima levam a modelos que permitem o estudo do desempe-

² Diz-se de um sistema ser invertível se distintas entradas levam a distintas saídas [11]. ³ Fenômeno que acontece quando os coeficientes do filtro LMS são atualizados muito lentamente devido ao excesso de erro de quantização [12].

nho do algoritmo LMS quantizado para amplitudes de sinais relativamente pequenas. Em aplicações práticas, os sinais envolvidos podem atingir amplitudes comparáveis com os limites de saturação dos registros internos do DSP. Nestes casos, um modelo linear ou não-linear aproximado para o erro de quantização pode não representar fielmente o comportamento do algoritmo LMS quando implementado digitalmente. Um modelo linear é satisfatório durante os primeiros estágios de adaptação, quando o erro é grande, e caso não exista saturação. Quando o erro decresce (regime estacionário) ou quando ocorrem saturações, o comportamento do algoritmo não pode ser previsto por um modelo linear [8] ou por um modelo não-linear sem saturação.

Assim, motivada por este interesse prático, neste trabalho analisa-se algumas propriedades do comportamento do algoritmo LMS em regime permanente, considerando sinais e coeficientes representados em precisão finita, incluindo os efeitos da saturação.

A seguir apresentam-se os objetivos desta dissertação.

1.4 METODOLOGIA UTILIZADA

Levando em consideração a importância do estudo dos efeitos da precisão finita na filtragem adaptativa, descrito na seção 1.3, este trabalho propõe uma nova modelagem do algoritmo LMS, considerando sinais e coeficientes em precisão finita e incluindo os efeitos de saturação. Estuda-se o comportamento dos coeficientes do filtro adaptativo em precisão finita, modelando-os como uma cadeia de Markov de primeira ordem. Devido à grande complexidade matemática, considera-se apenas o caso unidimensional (um coeficiente). Como as probabilidades de transição são desconhecidas, faz-se uma análise estatística determinando a função densidade de probabilidade (fdp) do coeficiente do filtro. Conhecidas essas probabilidades e aplicando conhecimentos das cadeias de Markov, estabelece-se as propriedades a serem satisfeitas pela matriz de probabilidades de transição para que o algoritmo LMS em precisão finita convirja para o coeficiente ótimo. Finalmente, determina-se a condição do passo de adaptação para satisfazer essas propriedades: primeiro, faz-se uma

análise para o caso com ruído muito baixo. Em seguida, estuda-se o comportamento do LMS em precisão finita no caso de ruído elevado.

1.5 DIVISÃO DE CAPÍTULOS

No Capítulo 2, faz-se uma revisão de alguns conceitos referentes a variáveis aleatórias com o objetivo de compreender melhor a determinação da fdp do coeficiente do filtro adaptativo. Também faz-se uma revisão de cadeias de Markov de primeira ordem, em que analisam-se tópicos como tipos de cadeias, propriedades de classes de estados, probabilidades de estado estacionário e outros. Importantes conclusões são tiradas a respeito das propriedades a serem satisfeitas pela matriz de probabilidades de transição para que o algoritmo LMS em precisão finita convirja para o coeficiente ótimo.

No Capítulo 3, apresenta-se o desenvolvimento matemático para obtenção da fdp do coeficiente adaptativo. Simulações de Monte Carlo comprovam a qualidade do modelo proposto.

No Capítulo 4, conhecidas as probabilidades de transição entre os estados dos coeficientes e as propriedades da matriz de probabilidades, determinam-se as condições do passo de adaptação para a convergência do algoritmo para o coeficiente de Wiener. Primeiro, faz-se uma análise para o caso com ruído muito baixo. Em seguida, determina-se o tamanho do passo em função da potência do ruído. Importantes conclusões são tiradas a respeito do comportamento do algoritmo operando em situações reais.

Finalmente, no Capítulo 5 apresentam-se as conclusões mais relevantes deste trabalho e sugestões para trabalhos futuros.

1.6 OBJETIVOS DO TRABALHO

1.6.1 Objetivo geral

Este trabalho tem como objetivo geral analisar o comportamento do algoritmo LMS em regime permanente, considerando sinais e coeficientes representados em precisão finita e in-

cluindo os efeitos de saturação.

O estudo é feito para o caso de dimensão 1 (um único coeficiente no filtro adaptativo) para evitar a complexidade algébrica inerente ao caso multidimensional. O estudo deste caso, no entanto, permite a obtenção de informações importantes sobre o comportamento do algoritmo, que não poderiam ser obtidas usando os modelos existentes na literatura. O coeficiente adaptativo é modelado como uma cadeia de Markov. A sua função densidade de probabilidade (fdp) é determinada, condicionada ao valor do coeficiente anterior e ao estado do sinal de entrada. A partir da fdp dos coeficientes, determina-se a matriz de probabilidades de transição da cadeia de Markov e, através de suas propriedades, estuda-se a convergência e o comportamento do algoritmo em regime permanente. Os objetivos específicos para este trabalho são:

1.6.2 Objetivos específicos

- Modelar os coeficientes do filtro adaptativo como estados de uma cadeia de Markov, para o caso unidimensional, determinando a função densidade de probabilidade (fdp) dos coeficientes condicionada ao coeficiente anterior e ao estado do sinal de entrada.
- Estabelecer as propriedades a serem satisfeitas pela matriz de probabilidade de transição P_w para que o algoritmo convirja para o coeficiente ótimo.
- Determinar condições sobre o passo de adaptação para a que o algoritmo convirja para a solução de Wiener.
- Validar, mediante simulações de Monte Carlo, o modelo proposto e as condições de passo de adaptação.

1.7 CONTRIBUIÇÕES DO TRABALHO

Da seção 1.3.4, revisão bibliográfica, tem-se que análise feitas sobre o comportamento do LMS em precisão finita, tais como [4] e [5] utilizam um modelo linearizado branco não

correlacionado para o erro de quantização. No entanto, sendo a operação do quantizador não linear, o desempenho do LMS quantizado não pode ser previsto por uma análise linear. Após [3], [7]-[9] incorporam o efeito não linear no comportamento dos coeficientes obtendo uma curva de aprendizado recursiva mais exata para o EMQ, mas este modelo só permite o estudo do desempenho do LMS quantizado para amplitudes de sinais relativamente pequena, já que não são considerados os efeitos de saturação. O sinal de entrada é considerado em precisão infinita na maioria dos modelos.

Nesta dissertação propõem-se uma nova modelagem do algoritmo LMS operando em precisão finita, levando em conta todas as não-linearidades: os efeitos de saturação e os efeitos de quantização nos sinais de entradas e os coeficientes, tendo como proposta inédita o modelamento dos coeficientes do filtro como uma cadeia de Markov. Desta forma obtém-se informações não existentes na literatura sobre o comportamento do LMS quando implementado em precisão finita.

Capítulo 2

REVISÃO DE VARIÁVEL ALEATÓRIA E CADEIAS DE MARKOV

2.1 INTRODUÇÃO

Um dos objetivos desta dissertação é determinar as probabilidades de ocorrência dos estados dos coeficientes do filtro para o algoritmo LMS mediante a determinação da função densidade de $w(n)$. Assim, para uma melhor compreensão deste estudo, este capítulo apresenta uma breve descrição dos principais conceitos empregados nos próximos capítulos. Em geral, este capítulo considera predominantemente variáveis aleatórias discretas, já que somente o ruído de medição $z(n)$ é uma variável aleatória contínua.

Em seguida, apresenta-se uma breve revisão de cadeias de Markov. As cadeias de Markov são ferramentas que permitem analisar o comportamento de determinados tipos de processos estocásticos que evoluem de forma não determinística com o passar do tempo em torno de um conjunto de estados. As cadeias de Markov podem modelar o comportamento de sistemas que variam com o passar do tempo de forma que cada variação do sistema dependa de estados anteriores deste mesmo sistema. Então, é possível estudar o comportamento do algoritmo adaptativo LMS quando implementado em aritmética de precisão finita mediante a modelagem dos coeficientes do filtro como estados numa cadeia de Markov.

2.2 VARIÁVEL ALEATÓRIA

2.2.1 Definição de variável aleatória

Define-se uma variável aleatória real $X(\xi)$ como um mapeamento que atribui um número real x a cada evento ξ de um espaço abstrato de possibilidades S .

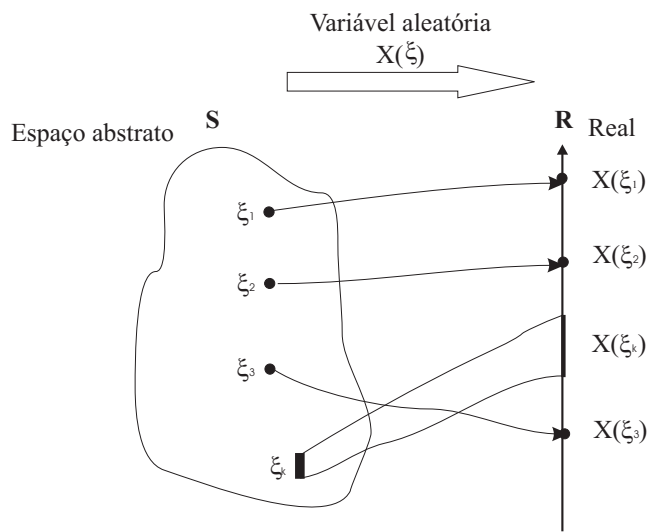


Figura 2.1: Ilustração gráfica do mapeamento de uma variável aleatória [2]

Assim,

$$X(\xi) = x$$

onde $X(\xi)$ é a função variável aleatória e x é o valor da variável aleatória.

Este mapeamento deve satisfazer três condições: (1) o intervalo $X(\xi) \leq x$ é um evento no espaço abstrato de probabilidade para cada x ; (2) $\Pr\{X(\xi) = \infty\} = 0$; e (3) $\Pr\{X(\xi) = -\infty\} = 0$ [2].

Observe os exemplos a seguir:

Exemplo 2.1.1: Para o lançamento de um dado tem-se os resultados possíveis $S = \{1, 2, 3, 4, 5, 6\}$,

onde alguns dos possíveis eventos podem ser (Figura 2.1):

ξ_1 : ocorrência de qualquer número, $\xi_1 = \{1, 2, 3, 4, 5, 6\}$

ξ_2 : ocorrência de um número ímpar, $\xi_2 = \{1, 3, 5\}$

ξ_3 : ocorrência de um número par, $\xi_3 = \{2, 4, 6\}$

Assim, as probabilidades de ocorrência dos diferentes eventos são dados por $\Pr\{\xi = \xi_1\}$, $\Pr\{\xi = \xi_2\}$ e $\Pr\{\xi = \xi_3\}$.

Exemplo 2.1.2: De [13], a Figura 2.2 ilustra uma roda da fortuna, onde os possíveis resultados do ponteiro ao girar são os números de 0 até 12, indicados na roda. Assim o espaço abstrato de possibilidades S , contém todos os números no intervalo $0 < s \leq 12$. Então defina-se a variável aleatória pela função,

$$X = X(s) = s^2$$

assim os pontos de S são mapeados no intervalo $0 < x \leq 144$ da linha real.

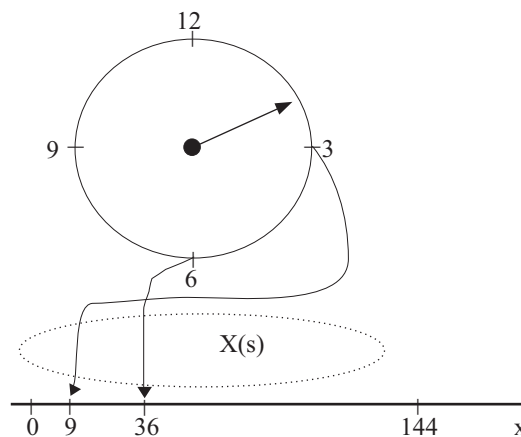


Figura 2.2: Ilustração gráfica do mapeamento de uma variável aleatória para o Exemplo 2.1.2 [13]

Uma *variável aleatória é discreta* se pode assumir um número finito ou infinito numerável de valores em \mathbf{R} . O espaço abstrato para uma variável aleatória discreta pode ser

discreto, contínuo ou uma mistura de ambos. Por exemplo, para o experimento da roda da fortuna, pode definir-se uma variável aleatória discreta a partir do espaço abstrato S , de valores contínuos, usando uma função cujo valor seja igual a 1, para $0 < s \leq 6$ e -1 para $6 < s \leq 12$. Assim, a variável aleatória é discreta e está definida em um espaço contínuo.

Uma *variável aleatória é contínua* se pode assumir qualquer valor real compreendido dentro de um intervalo. Não pode ser produzida a partir de um espaço abstrato discreto porque todas as variáveis aleatórias são produto de uma única função para todos os pontos do espaço abstrato. Assim, uma variável puramente contínua não pode resultar de um espaço misto devido à presença da parte discreta.

2.2.2 Função Distribuição e Função Densidade

A expressão $P\{X \leq x\}$ é a probabilidade do evento $\{X \leq x\}$, assim $P\{X \leq x\}$ é função de x . Esta função denota-se por $F_X(x)$, *função de distribuição de probabilidade cumulativa* (ou simplesmente função distribuição) da variável aleatória X [13],

$$F_X(x) = P\{X \leq x\} \quad (2.1)$$

Existem propriedades específicas da função distribuição derivadas do fato que $F_X(x)$ é uma probabilidade [13]:

1. $F_X(-\infty) = 0$
2. $F_X(\infty) = 1$
3. $0 \leq F_X(x) \leq 1$
4. $F_X(x_1) \leq F_X(x_2)$ se $x_1 < x_2$
5. $P\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1)$
6. $F_X(x^+) = F_X(x)$

Se X é uma variável aleatória discreta, considerando a definição de função distribuição (2.1), então $F_X(x)$ deve ter a forma de uma escada como mostra a Figura 2.3. A amplitude de cada degrau nesta escada é igual à probabilidade de ocorrência de um valor $X = x$, em que o degrau ocorre. Se os valores possíveis de X são denotados x_i e $i = 1, \dots, N$, [13]:

$$F_X(x) = \sum_{i=1}^N P\{X = x_i\} u\{x = x_i^-\} \quad (2.2)$$

em que $u(\cdot)$ é a função degrau definida por:

$$u(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

Introduzindo a notação

$$P(x_i) = P\{X = x_i\}$$

então pode escrever-se a equação (2.2) como:

$$F_X(x) = \sum_{i=1}^N P\{x_i\} u\{x = x_i^-\} \quad (2.3)$$

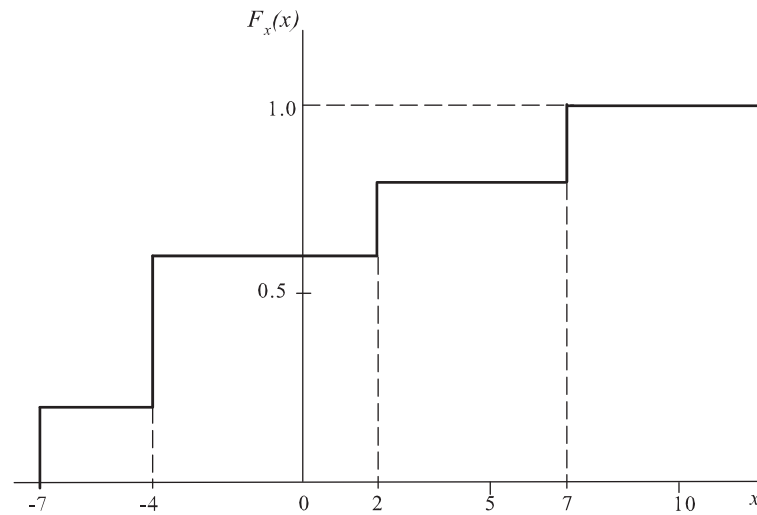
A seguir apresenta-se um exemplo de $F_X(x)$ para variáveis aleatórias discretas.

Exemplo 2.1.3: Considere X com valores discretos no conjunto $\{-7, -4, 2, 7\}$ com probabilidades de ocorrência dadas, respectivamente, por $\{0,2, 0,4, 0,2, 0,2\}$. A Figura 2.3a apresenta o gráfico de $F_X(x)$.

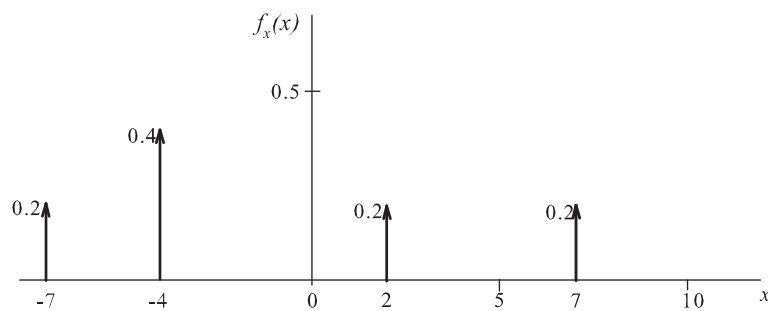
Observe que $P\{X < -7\} = 0$ já que não existem pontos no espaço abstrato para $X < -7$ (Propriedade 1). Só quando $X = -7$ obtém-se um salto em $F_X(x)$ com probabilidade 0,2.

Para os valores $-7 < x < -4$ a função distribuição permanece constante com valor igual a 0,2. Quando $x = -4$ então $F_X(x)$ dá um salto de 0,4. Assim o processo continua até que

todos os pontos do espaço abstrato são considerados e $F_X(x) = 1$.



a)



b)

Figura 2.3: Função distribuição (a) e função densidade (b) (baseado em [13])

A Figura 2.3b ilustra a função densidade do *Exemplo 3.1.3*. A *função densidade* $f_X(x)$ corresponde à derivada da função distribuição:

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.4)$$

Existência de $f_X(x)$: Para uma variável aleatória discreta com função de distribuição dada por (2.3), a diferenciação de $F_X(x)$ resulta em [13],:

$$f_X(x) = \sum_{i=1}^N P(x_i) \delta(x - x_i^-) \quad (2.5)$$

A função densidade para uma variável aleatória discreta existe no sentido que utiliza-se a função impulso para descrever a derivada de $F_X(x)$ em cada ponto de descontinuidade. Por exemplo a Figura 2.3b apresenta a função densidade correspondente à função distribuição da Figura 2.3a.

Propriedades: Algumas propriedades que $f_X(x)$ deve satisfazer são:

1. $0 \leq f_X(x)$
2. $\int_{-\infty}^{\infty} f_X(x) dx = 1$
3. $F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi$
4. $P\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} f_X(x) dx$

As propriedades 1 e 2 requerem que a função densidade seja não- negativa e tenha uma área unitária. Essas duas propriedades devem ser usadas para testar se uma função de densidade é válida. A propriedade 3 é outra forma de escrever (2.4). A propriedade 4 relaciona a probabilidade de X estar compreendida entre x_1 e x_2 (incluso) e sua função densidade.

2.2.3 A variável aleatória normal

Uma variável aleatória é de distribuição normal de média zero e variância σ^2 , ($N(0, \sigma^2)$), se sua função densidade é dada por uma exponencial na forma:

$$f_X(x) = A e^{-\frac{1}{2\sigma^2} x^2} \quad (2.6)$$

Sabe-se que:

$$\int_{-\infty}^{\infty} e^{-\alpha z^2} dz = \sqrt{\frac{\pi}{\alpha}} \quad (2.7)$$

e, por propriedade da função distribuição que,

$$\int_{-\infty}^{\infty} f(x) dx = F(\infty) - F(-\infty) = 1 \quad (2.8)$$

Usando (2.6) em (2.8) e considerando (2.7) tem-se,

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= A \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} x^2} dx = A \sqrt{2\pi\sigma^2} = 1 \\ \text{logo} \quad A &= \frac{1}{\sqrt{2\pi\sigma^2}} \end{aligned} \quad (2.9)$$

Assim, (2.6) resulta,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (2.10)$$

com uma mudança na variável x obtém-se a função densidade normal para x com média η , $N(\eta, \sigma^2)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\eta)^2}{2\sigma^2}} \quad (2.11)$$

em que $\sigma > 0$ e $-\infty < \eta < \infty$ são constantes reais.

A Figura 2.4, apresenta uma função gaussiana $N(\eta, \sigma^2)$. Note que o máximo ponto $(2\pi\sigma^2)^{-\frac{1}{2}}$ ocorre para $x = \eta$.

A densidade gaussiana ou normal, como também é conhecida, é a mais importante das densidades contínuas, pois muitas variáveis aleatórias de ocorrência natural ou de processos práticos obedecem a esta função densidade. Por exemplo a tensão nos terminais de uma resistência na saída de um amplificador pode ser aleatória devido a uma corrente aleatória,

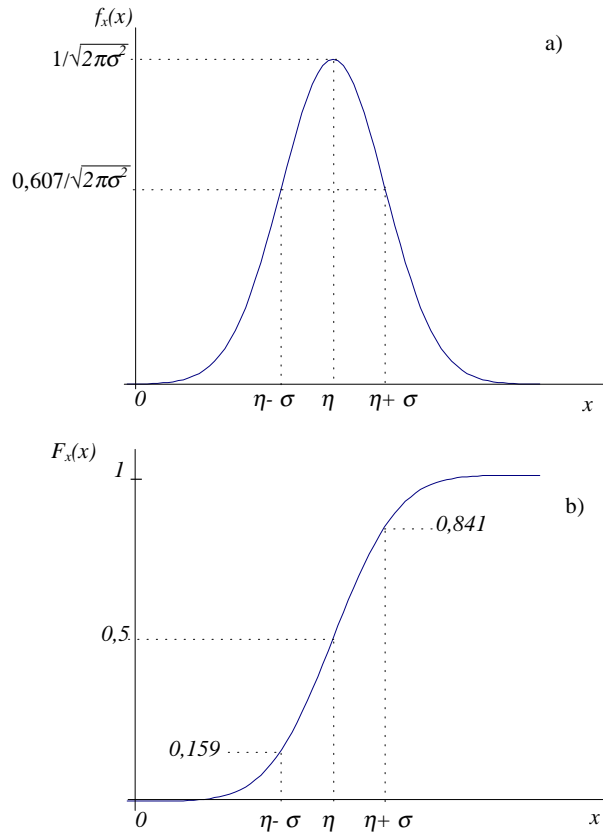


Figura 2.4: Função densidade (a) e função distribuição (b) para uma variável aleatória gaussiana.

resultado de muitas contribuições de outras correntes aleatórias circulantes dentro do amplificador. Este tipo de ruído é conhecido como *gaussiano* porque a variável aleatória que representa a tensão correspondente tem uma função de densidade gaussiana [13].

A partir da propriedade 3, para uma função densidade $N(\eta, \sigma^2)$ tem-se que:

$$F_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(x-\eta)^2}{2\sigma^2}} dx \quad (2.12)$$

Esta integral não tem uma solução na forma fechada e deve ser avaliada numericamente ou por métodos aproximados. Assim considerando que, sua integral no intervalo $(0, x)$ é dada

por

$$\begin{aligned} erf(x) &= A \int_0^x e^{-\frac{z^2}{2}} dz \\ erf(-x) &= -erf(x) \\ erf(\infty) &= A \int_0^\infty e^{-\frac{z^2}{2}} dz = \frac{1}{2} \end{aligned}$$

então

$$\begin{aligned} A \int_{-\infty}^x e^{-\frac{1}{2} \frac{(z-\eta)^2}{\sigma^2}} dz &= A \int_{-\infty}^0 e^{-\frac{1}{2} \frac{(z-\eta)^2}{\sigma^2}} dz + A \int_0^x e^{-\frac{1}{2} \frac{(z-\eta)^2}{\sigma^2}} dz \\ &= -A \int_0^{-\infty} e^{-\frac{1}{2} \frac{(z-\eta)^2}{\sigma^2}} dz + erf\left(\frac{x-\eta}{\sigma}\right) \\ &= \frac{1}{2} + erf\left(\frac{x-\eta}{\sigma}\right) \end{aligned} \quad (2.13)$$

e,

$$F_x(x) = \frac{1}{2} + erf\left(\frac{x-\eta}{\sigma}\right) \quad (2.14)$$

O valor da função $erf(x)$ encontra-se por tabela. (Ver Apêndice A)

A partir da propriedade 5 para a função distribuição tem-se que:

$$P\{x_1 < x \leq x_2\} = F_X(x_2) - F_X(x_1) = erf\left(\frac{x_2-\eta}{\sigma}\right) - erf\left(\frac{x_1-\eta}{\sigma}\right)$$

Exemplo 2.1.4: Considere uma variável aleatória distribuída normalmente com $\eta = 1000$, $\sigma = 50$. Determinar a probabilidade de que x esteja entre 900 e 1050. Aplicando o resultado anterior,

$$\begin{aligned} P\{900 < x \leq 1050\} &= erf\left(\frac{1050-1000}{50}\right) - erf\left(\frac{900-1000}{50}\right) \\ &= erf(1) - erf(-2) = 0,819 \end{aligned}$$

2.2.4 Outras distribuições e densidades

Discreta Uniforme: Um experimento diz-se que segue uma distribuição discreta uniforme se apresenta $\{1, 2, \dots, N\}$ resultados distintos, todos eles equiprováveis e tal que a probabilidade de cada resultado venha dada mediante a seguinte função de probabilidade:

$$P\{X = x\} = P\{x\} = \frac{1}{N} \quad \text{dado que } x = 1, 2, 3, \dots, N$$

Assim, das equações (2.3) e (2.5) tem-se que:

$$F_x(x) = \frac{1}{N} \sum_{i=1}^N u(x - x_i^-) \quad (2.15)$$

$$f_x(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i) \quad (2.16)$$

Binomial: É uma distribuição que ocorre em problemas de sistemas de radar e sonar, assim como em experimentos que tem só dois possíveis resultados em um dado processo.

Considerando $0 < p < 1$ e $N = 1, 2, \dots$ a função,

$$f_x(x) = \sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} \delta(x - k) \quad (2.17)$$

denomina-se função densidade binomial. A expressão $\binom{N}{k}$ é o coeficiente binomial dado por:

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

Integrando a equação (2.17), tem-se a função distribuição,

$$F_x(x) = \sum_{k=0}^N p^k (1-p)^{N-k} u(x - k^-) \quad (2.18)$$

Poisson: A variável aleatória de Poisson é aplicável a fenômenos nos quais eventos aleatórios ocorrem através do tempo tais como acidentes, ligações falsas em sistemas telefônicos, erros tipográficos nas páginas de um livro, etc. As funções densidade e distribuição são dadas por:

$$f_x(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} \delta(x - k) \quad (2.19)$$

$$F_x(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} u(x - k^-) \quad (2.20)$$

2.2.5 Distribuição e densidade da soma de variáveis aleatórias

Nesta seção descreve-se a determinação da função distribuição e densidade da soma de variáveis aleatórias estatisticamente independentes. Inicialmente, enuncia-se algumas propriedades da função densidade conjunta de duas variáveis aleatórias.

Propriedades [13]

1. $f_{\mathbf{x},\mathbf{y}}(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dx dy = 1$
3. $F_{\mathbf{x},\mathbf{y}}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{\mathbf{x},\mathbf{y}}(\xi_1, \xi_2) d\xi_1 d\xi_2$
4. $F_{\mathbf{x}}(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(\xi_1, \xi_2) d\xi_2 d\xi_1$
 $F_{\mathbf{y}}(y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(\xi_1, \xi_2) d\xi_1 d\xi_2$
5. $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{\mathbf{x},\mathbf{y}}(x, y) dx dy$
6. $f_{\mathbf{x}}(x) = \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dy$
7. $f_{\mathbf{y}}(y) = \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dx$

Considere W como o resultado da soma de duas variáveis aleatórias, X e Y

$$W = X + Y$$

Então a distribuição de probabilidade de W é definida por,

$$F_w(w) = P\{W \leq w\} = P\{X + Y \leq w\}$$

A Figura 2.5 ilustra o plano xy , onde $x + y \leq w$ é a metade do plano à esquerda da linha

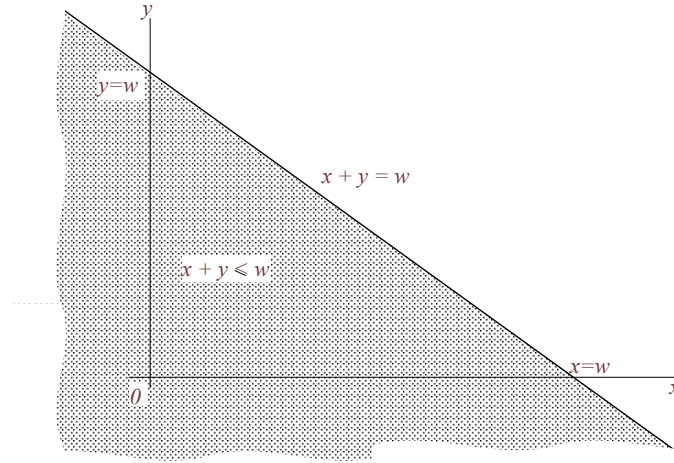


Figura 2.5: Região no plano xy onde $x + y \leq w$ [13]

$x + y = w$. Da propriedade 5 antes descrita, temos que a função distribuição de probabilidade de W é dada por:

$$F_w(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{w-y} f_{x,y}(x,y) dx dy \quad (2.21)$$

Diferenciando (2.21) com respeito a w tem-se:

$$f_w(w) = \int_{-\infty}^{\infty} f_{x,y}(w-y,y) dy \quad (2.22)$$

Quando X e Y são variáveis independentes, a função densidade da soma $X + Y$ reduz-se à *convolução* de suas respectivas densidades, assim

$$f_w(w) = \int_{-\infty}^{\infty} f_x(w-y)f_y(y) dy \quad (2.23)$$

2.2.6 Funções de distribuição e densidade condicionadas

Distribuição condicional: Para dois eventos A e B onde $P\{B\} \neq 0$, a probabilidade condicional de A dado que B tenha ocorrido é,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Considere A como o evento $\{X \leq x\}$ para a variável aleatória X . A probabilidade resultante $P\{X \leq x|B\}$ define-se como a *função distribuição condicional* de X e denota-se por;

$$F_x(x|B) = P\{X \leq x|B\} = \frac{P\{X \leq x \cap B\}}{P(B)}$$

A notação $\{X \leq x \cap B\}$ implica o evento conjunto $\{X \leq x\} \cap B$. Este evento conjunto consiste de todos os resultados tais que:

$$X(s) \leq x \quad \text{e} \quad s \in B$$

A distribuição condicional aplica-se a variáveis aleatórias discretas, contínuas e mistas.

Propriedades da função distribuição condicional [13]

1. $F_X(-\infty|B) = 0$
2. $F_X(\infty|B) = 1$
3. $0 \leq F_X(x|B) \leq 1$
4. $F_X(x_1|B) \leq F_X(x_2|B) \quad x_1 < x_2$
5. $P\{x_1 < X \leq x_2|B\} = F_X(x_2|B) - F_X(x_1|B)$
6. $F_X(x^+|B) = F_X(x|B)$

Observe que estas propriedades têm o mesmo significado que as propriedades descritas para a função distribuição ordinária.

Densidade Condicional: De maneira similar à função densidade ordinária, define-se a *função densidade condicional* da variável aleatória X como a derivada da função distribuição condicional. Assim,

$$f_X(x|B) = \frac{dF_X(x|B)}{dx}$$

se $F_X(x|B)$ apresenta descontinuidades como acontece com as variáveis aleatórias discretas, então a função impulso aparecerá em $f_X(x|B)$.

Como a densidade condicional está relacionada com a distribuição condicional mediante a derivada desta última, satisfaz as mesmas propriedades da função densidade ordinária.

Propriedades da função densidade condicional [13]

1. $f_X(x|B) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x|B)dx = 1$
3. $F_X(x|B) = \int_{-\infty}^x f_X(\xi|B)d\xi$
4. $P\{x_1 < X \leq x_2|B\} = \int_{x_1}^{x_2} f_X(x|B)dx$

2.2.7 Valor esperado de uma variável aleatória

O valor esperado (ou simplesmente esperança) de uma variável aleatória X , é a soma da probabilidade de ocorrência de cada evento multiplicado por seu valor, ou seja,

$$E[X] = \bar{X} = \int_{-\infty}^{\infty} x f_X(x)dx \quad (2.24)$$

Considerando que $f_X(x)$ para uma variável aleatória discreta é dada por

$$f_X(x) = \sum_{i=1}^N P(x_i) \delta(x - x_i) \quad (2.25)$$

e substituindo a equação (2.25) na (2.24) tem-se que o valor esperado de uma variável aleatória discreta é dado por

$$E[X] = \sum_{i=1}^N x_i P(x_i) \quad (2.26)$$

Momentos

De acordo com o desenvolvido nesta dissertação, só se apresentará o momento de ordem 2, denominado variância.

$$var[X] = \sigma_x^2 = E[(X - \bar{X})^2] = \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x) dx \quad (2.27)$$

A raiz quadrada da variância σ_x^2 , denomina-se desvio padrão de X e é uma medida da dispersão na função $f_X(x)$ em relação à média.

A variância de uma variável aleatória discreta é obtida substituindo (2.25) em (2.27)

$$var[X] = \sigma_x^2 = \sum_{i=1}^N (x_i - \bar{X})^2 P(x_i) \quad (2.28)$$

2.3 CADEIAS DE MARKOV

2.3.1 Definição e propriedades básicas

Uma cadeia de Markov é uma seqüência (enumerável) de variáveis aleatórias $X(n)$, definidas em um espaço de estados discretos que obedecem à propriedade de Markov. [15]

Sabendo que no algoritmo LMS o estado futuro de $w(n)$ só depende do estado presente de $w(n)$, analisa-se neste capítulo o caso particular de uma cadeia de Markov de ordem 1, em que o estado futuro j , só depende do estado presente i , ou seja:

A probabilidade P_{ij} de $X(n+1)$ estar no estado futuro j depende apenas do estado presente i e não dos estados passados. Isto é:

$$\begin{aligned} P_{ij} &= Pr\{X(n+1) = j | X(0) = k_0, X(1) = k_1, \dots, X(n-1) = k_{n-1}, X(n) = i\} \\ &= Pr\{X(n+1) = j | X(n) = i\}, \quad \forall n = 0, 1, 2, 3, \dots \quad \forall k_0, k_1, \dots, k_{n-1}, i, j \in X \end{aligned} \quad (2.29)$$

Se a probabilidade de ocorrência de cada estado na equação (2.29) é independente de n , ou seja é constante ao longo do tempo, a cadeia de Markov denomina-se de homogênea, (HMC, Homogeneous Markov Chain) [14].

Em (2.29) k_0, k_1, \dots, k_{n-1} , i e j denominam-se *estados* e correspondem a números inteiros não negativos, e n representa o índice de tempo. Se $n \in \mathbf{Z}^+$ então a cadeia de Markov denomina-se de tempo discreto. Por exemplo, se $X(n) = i$ diz-se que o *processo está no estado i no momento n*

Uma cadeia de Markov fica completamente definida quando os estados de X pertencentes a $\{0, 1, 2, \dots, N\}$ e as probabilidades de transição entre estados em um período são conhecidos.

Os *estados* são uma caracterização da situação em que se encontra o sistema em um dado instante. Essa caracterização pode ser qualitativa ou quantitativa. O estado de um sistema em um instante n é uma variável cujo valor só pode pertencer ao conjunto de estados possíveis do sistema. A mudança de estados do sistema denomina-se *transição*.

Por ser o sistema estocástico, o seu estado em um determinado instante não é conhecido com certeza. Conhece-se apenas a probabilidade associada a cada um dos estados. Na equação (2.29), P_{ij} recebe o nome de probabilidade de transição do estado i ao estado j .

2.3.2 Matriz de probabilidade de transição

Em uma cadeia de Markov finita e de ordem 1 uma forma conveniente de expressar o conjunto de probabilidades condicionais é mediante a matriz de probabilidade de transição \mathbf{P} [15]. \mathbf{P} é quadrada, com tantas linhas e colunas quanto o número de estados do sistema. O elemento (i, j) da matriz é igual à probabilidade de que j seja o próximo estado se i é o estado presente.

Como o sistema deve evoluir em n para algum dos N estados possíveis, as probabilidades de transição possuem a seguinte propriedade:

$$\sum_{j=1}^n p_{ij} = 1 \quad \text{para todos os estados } i. \quad (2.30)$$

Além disso, por definição cada uma das probabilidades deve ser não negativa:

$$p_{ij} \geq 0 \quad \text{para todos os estados } i, j \quad (2.31)$$

Assim, a soma dos elementos de cada linha da matriz \mathbf{P} é sempre igual a 1.

A matriz \mathbf{P} que satisfaz as propriedades descritas nas equações (2.29), (2.30), (2.31) denomina-se *matriz estocástica* [14].

Veja o seguinte exemplo, onde $X = \{0, 1, 2, 3, 4\}$ indica os estados de transição da cadeia. Ilustra-se X no vetor linha acima de \mathbf{P} e no vetor coluna do lado de \mathbf{P} .

$$\begin{array}{c} \text{Estados} \end{array} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \mathbf{P} = \begin{pmatrix} 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,3 & 0,4 & 0 & 0,1 \\ 0,1 & 0,3 & 0,2 & 0,3 & 0,1 \\ 0,6 & 0,1 & 0 & 0,1 & 0,2 \\ 0,3 & 0,2 & 0,2 & 0,2 & 0,1 \end{pmatrix}$$

Observe que $p_{34} = 0,2$ significa que a probabilidade de transição para ir do *estado* 3 ao *estado* 4 é igual a 0,2 (20%). Observe também como a soma dos elementos de cada linha (1 até 5) é sempre 1.

2.3.3 Equação de Chapman-Kolmogorov

Considerando que as probabilidades de transição são independentes de n , cadeia de Markov Homogênea, então é interessante conhecer as propriedades de transição depois de n passos, definidas por:

$$p_{ij}^n = Pr\{X_n = j \mid X_0 = i\}, \quad n \geq 0, \quad i, j \geq 0$$

Esta é a probabilidade de o processo encontrar-se no estado j se n períodos antes se encontrava no estado i .

As equações de Chapman-Kolmogorov fornecem um método para cálculo das probabilidades de transição em n períodos. Estas equações são dadas por [15]:

$$p_{ij}^{n+m} = \sum_{k=0}^{\infty} p_{ik}^n p_{kj}^m, \quad \forall n, m \geq 0, \quad \forall i, j \quad (2.32)$$

Na equação (2.32), p_{ik}^n é a probabilidade de que o sistema vá do estado i ao estado k em n passos. Assim, $p_{ik}^n p_{kj}^m$ representa a probabilidade de que o sistema vá do estado i ao estado j em $n + m$ passos, passando pelo estado k no passo n . Portanto, somando as probabilidades de todos os passos intermediários k obtemos a probabilidade de que o processo esteja em j depois de $n + m$ passos.

Formalmente tem-se,

$$\begin{aligned}
 p_{ij}^{n+m} &= Pr\{X_{n+m} = j \mid X_0 = i\} \\
 &= Pr\{X_{n+m} = j, X_n = k \mid X_0 = i\} \\
 &= Pr\{X_{n+m} = j, \mid X_n = k, X_0 = i\} Pr\{X_n = k \mid X_0 = i\} \\
 &= \sum_{k=0}^{\infty} p_{ik}^n p_{kj}^m
 \end{aligned}$$

Da equação (2.32), também pode dizer-se que:

$$p_{ij}^m = \sum_{k=0}^{\infty} p_{kj}^{m-1} p_{ik} \quad (2.33)$$

Da equação (2.33) pode obter-se a matriz $\mathbf{P}^{(m)}$ de transição de m passos a partir das potências da matriz \mathbf{P} ,

$$\begin{aligned}
 \mathbf{P}^{(2)} &= \mathbf{P}^{(1)} \cdot \mathbf{P}^{(1)} = \mathbf{P}^2 \\
 \mathbf{P}^{(3)} &= \mathbf{P}^{(2)} \cdot \mathbf{P}^{(1)} = \mathbf{P}^{(1)} \cdot \mathbf{P}^{(2)} = \mathbf{P}^2 \cdot \mathbf{P} = \mathbf{P}^3 \\
 \mathbf{P}^{(m)} &= \mathbf{P}^{(m-1)} \cdot \mathbf{P} = \mathbf{P} \cdot \mathbf{P}^{(m-1)} = \mathbf{P}^{m-1} \cdot \mathbf{P} = \mathbf{P}^m, \quad m = 0, 1, \dots
 \end{aligned} \quad (2.34)$$

em que $\mathbf{P}^{(1)} = \mathbf{P}$ e $\mathbf{P}^0 = \mathbf{I}$, esta última representando as probabilidades de que não haja qualquer transição [15].

2.3.4 Probabilidades estacionárias e tipos de cadeias de Markov

A probabilidade de que o sistema se encontre em um determinado estado após estar em operação por um tempo indefinidamente longo denomina-se *probabilidade estacionária*. O estudo das probabilidades estacionárias permite analisar o comportamento das cadeias de Markov em regime permanente.

Como \mathbf{P}^m define as probabilidades após m transições, o estudo das probabilidades esta-

estacionárias pode ser feito calculando-se potências elevadas de \mathbf{P} [15]. Conceitualmente, o regime permanente é obtido calculando-se $\lim_{m \rightarrow \infty} \mathbf{P}^m$

Cadeias regulares. As cadeias de Markov regulares são cadeias em que as probabilidades estacionárias não dependem do estado inicial. Por exemplo considere a matriz de probabilidade de transição P_1 tal que,

$$P_1 = \begin{pmatrix} 0,2 & 0,8 \\ 0,6 & 0,4 \end{pmatrix}$$

Onde calculando P_1^{15} obtém-se as probabilidades estacionárias. Observe como todas linhas de P_1^{15} são iguais, isto indica que as probabilidades estacionárias dos diferentes estados são independentes do estado inicial.

$$P_1^{15} = \begin{pmatrix} 0,4285 & 0,5714 \\ 0,4285 & 0,5714 \end{pmatrix}$$

Nota: Escolhe-se $m = 15$ como suficiente para que a matriz P não varie mais (regime estacionário), procedimento adoptado nos seguintes exemplos.

Cadeias não regulares. As cadeias de Markov não regulares são cadeias em que as probabilidades estacionárias dependem do estado inicial. Por exemplo considere a matriz de probabilidade de transição P_2 tal que,

$$P_2 = \begin{pmatrix} 0,5 & 0,4 & 0,1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0,2 & 0,8 \\ 0 & 0 & 0,6 & 0,4 \end{pmatrix}$$

Calculando P_2^{15} tem-se

$$P_2^{15} = \begin{pmatrix} 0,000 & 0,800 & 0,086 & 0,114 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0,429 & 0,571 \\ 0 & 0 & 0,429 & 0,571 \end{pmatrix}$$

Observe que P_2^{15} tem apenas duas linhas iguais. As outras duas linhas tendem para vetores diferentes. Neste caso, diz-se que as probabilidades estacionárias dependem do estado inicial da cadeia.

Cadeias ergódicas. Denomina-se cadeias ergódicas as cadeias regulares cujas probabilidades estacionárias são não nulas [15]. Por exemplo a matriz de probabilidade P_1 exposta acima. Observe como P_1^{15} tem todas as linhas iguais. As probabilidades de estado estável são independentes do estado de origem.

Cadeias quase ergódicas. As cadeias quase-ergódicas levam este nome porque apesar de que existem probabilidades estacionárias nulas [15], elas tem um comportamento em estado estacionário de uma cadeia regular.

Observe o seguinte exemplo:

$$P_3 = \begin{pmatrix} 0,1 & 0,7 & 0,2 & 0 \\ 0,6 & 0,3 & 0,1 & 0 \\ 0 & 0 & 0,5 & 0,5 \\ 0 & 0 & 0,9 & 0,1 \end{pmatrix}$$

Calculando P_3^{150} tem-se

$$P_3^{150} = \begin{pmatrix} 0 & 0 & 0,6429 & 0,3571 \\ 0 & 0 & 0,6429 & 0,3571 \\ 0 & 0 & 0,6429 & 0,3571 \\ 0 & 0 & 0,6429 & 0,3571 \end{pmatrix}$$

Observe como P_3^{150} tem algumas probabilidades estacionárias nulas, mas todas as linhas da matriz tendem para um mesmo vetor.

Cadeias não ergódicas. Nas cadeias não ergódicas as probabilidades estacionárias dependem do estado inicial da cadeia [15] e algumas são nulas. Observe a matriz de probabilidades P_2 exposta anteriormente.

Cadeias cíclicas. As cadeias cíclicas são aquelas que têm um comportamento cíclico, elas não convergem para um valor concreto [15]. Por exemplo observe a seguir, como a matriz P_4^m , tende para diferentes matrizes nas transições pares e ímpares.

$$P_4 = \begin{pmatrix} 0,5 & 0,4 & 0,1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0,2 & 0 & 0,8 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Para $m = 2k$ e $m_1 = 2k + 1$, com k muito grande tem-se,

$$P_4^m = \begin{pmatrix} 0 & 0,08 & 0,60 & 0,32 \\ 0 & 0,2 & 0 & 0,8 \\ 0 & 0 & 1 & 0 \\ 0 & 0,2 & 0 & 0,8 \end{pmatrix}$$

$$P_4^{m_1} = \begin{pmatrix} 0 & 0,2 & 0,4 & 0,48 \\ 0 & 0 & 1 & 0 \\ 0 & 0,2 & 0 & 0,8 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Neste caso particular, P_4 corresponde a uma cadeia de período 2. Observe que como a primeira coluna é sempre zero, o estado 1 não aparecerá nas probabilidades a longo prazo. Essa cadeia é não ergódica e não regular

Com a finalidade de analisar o comportamento da matriz de probabilidades de transição em estado estacionário, é preciso estabelecer uma metodologia para que seja possível a partir da matriz \mathbf{P} de um passo:

- Classificar a matriz em ergódica ou não ergódica, como também em regular, não regular ou cíclica.
- Determinar a matriz de probabilidade estacionária.

Uma maneira de fazer isso é utilizando propriedades relativas à relação entre estados da matriz \mathbf{P} . Estas propriedades permitem, por sua vez, definir classes formadas por subconjuntos de estados. Também podem definir-se as propriedades das classes. Para entender as propriedades das classes, precisa-se primeiro estudar as propriedades dos estados da cadeia de Markov.

Propriedades de estados

Dados dois estados de uma cadeia, pode-se estabelecer dois tipos de relação entre eles:

- Diz-se que o estado j é *atingível* a partir do estado i ($i \rightarrow j$), se e só se houver pelo menos um caminho de i para j .
- Dois estados i e j são *comunicantes* se e só se j for atingível a partir de i e i for atingível a partir de j , $i \leftrightarrow j$
- Existirá um *ciclo* dentro de uma cadeia de Markov se existir um caminho na cadeia que comunique o estado i consigo mesmo. O circuito associado a este ciclo se caracterizará pelo número mínimo de transições que necessitará o sistema para voltar para o estado

i , quando o processo foi iniciado nesse estado. Este número constituirá a duração do ciclo.

Assim, a existência de um circuito implica que todos os estados que o formam estão comunicados. Convém que qualquer estado esteja comunicado consigo mesmo, já que assim se pode retornar a ele em uma transição.

Para analisar as definições anteriores, utiliza-se a representação gráfica da matriz de probabilidades de transição, conhecida como diagrama de transição de estados.

Diagrama de Transição de Estados: Em um diagrama de transição, representa-se cada estado da cadeia por um nó e cada transição de probabilidade não nula pela relação entre nós. Assim, é possível observar todas as situações de relacionamento entre um estado i e um estado j . Por exemplo, seja P_y definida como:

$$P_y = \begin{pmatrix} 0,1 & 0,5 & 0,4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0,8 & 0 & 0,2 \\ 0 & 0 & 0,3 & 0,7 \end{pmatrix}$$

O diagrama de estados correspondente é apresentado na Figura 2.6

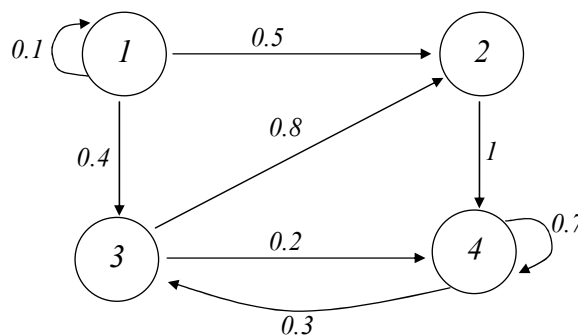


Figura 2.6: Diagrama de Transição de estados da matriz P_y (tomado de [15])

Note que os estados 3 e 4 são comunicantes. O estado 4 é atingível a partir do estado 1, mas o reverso não é possível. Observe também a existência dos ciclos $\{2, 4, 3, 2\}$, $\{3, 4, 3\}$,

$\{4, 4\}$ e $\{1, 1\}$.

Propriedade de classe

Dado que convencionou-se que um estado está sempre comunicado consigo mesmo, a relação de comunicação entre estados pode ser:

- a) $i \leftrightarrow i$ reflexiva
- b) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$ simétrica
- c) $i \leftrightarrow i, i \leftrightarrow j \Rightarrow i \leftrightarrow k$ transitiva

A relação de comunicação \leftrightarrow é denominada uma relação de equivalência. Estados ligados por uma relação de equivalência formam uma classe de equivalência. Assim, os estados de uma cadeia de Markov podem classificar-se em diversas classes. As classes de equivalência podem ser:

- *Classe final*. Se ao chegar a um dos estados da classe o processo evolui nas transições seguintes sempre dentro dos estados da classe.
- *Classes de passo*. Todas as classes que não sejam classes finais. Classes de passo têm um interesse limitado no estudo das cadeias de Markov.

No caso apresentado na Figura 2.7 a classe de equivalência $\{1, 2\}$ é uma classe de passo, enquanto que a classe de equivalência $\{3, 4\}$ é uma classe final.

Periodicidade de uma classe final: O período de uma classe define-se como o máximo divisor comum das durações dos ciclos encontrados em uma classe. Assim, existem dois tipos de classes finais:

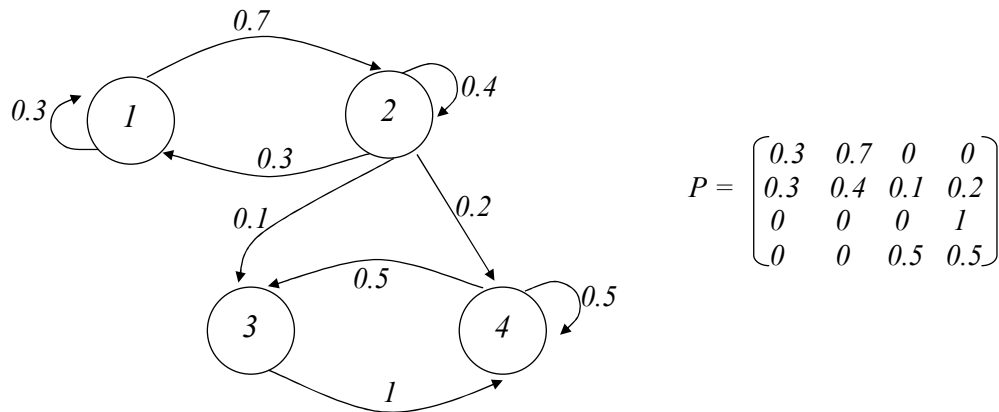


Figura 2.7: Propriedades de classes em uma cadeia de Markov (tomado de [15])

- Classes periódicas: Aquelas que têm um período de valor $p > 1$
- Classes aperiódicas: Aquelas que têm um período de valor $p = 1$.

A Figura 2.8 apresenta os gráficos de três classes finais de cadeias de Markov. Portanto, assume-se que nas respectivas cadeias não existem conexões saindo dessas classes. A classe da Figura 2.8a contém ciclos com duração iguais a 2 e 3. Portanto, como o máximo divisor comum é igual a 1, trata-se de uma classe aperiódica. Na Figura 2.8b observam-se ciclos com duração iguais a 2, 4 e 8. Assim, o período p é igual a 2 e essa classe é periódica. Na Figura 2.8c há um ciclo com duração igual a 1. Logo, $p = 1$ e a classe é aperiódica.

Com as propriedades definidas anteriormente, pode-se estabelecer duas classificações: segundo o número de classes finais e segundo a periodicidade de suas classes finais.

Classificação segundo o número de classes finais

Com relação às classes finais de uma cadeia de Markov pode-se ter as seguintes situações:

- Se a cadeia completa é ao mesmo tempo uma classe final, trata-se de uma *cadeia ergódica*. Cadeias deste tipo são denominadas cadeias irreduzíveis [16].
- Se a cadeia tem só uma classe final, mas tem classes de passo, trata-se de uma *cadeia quase-ergódica*

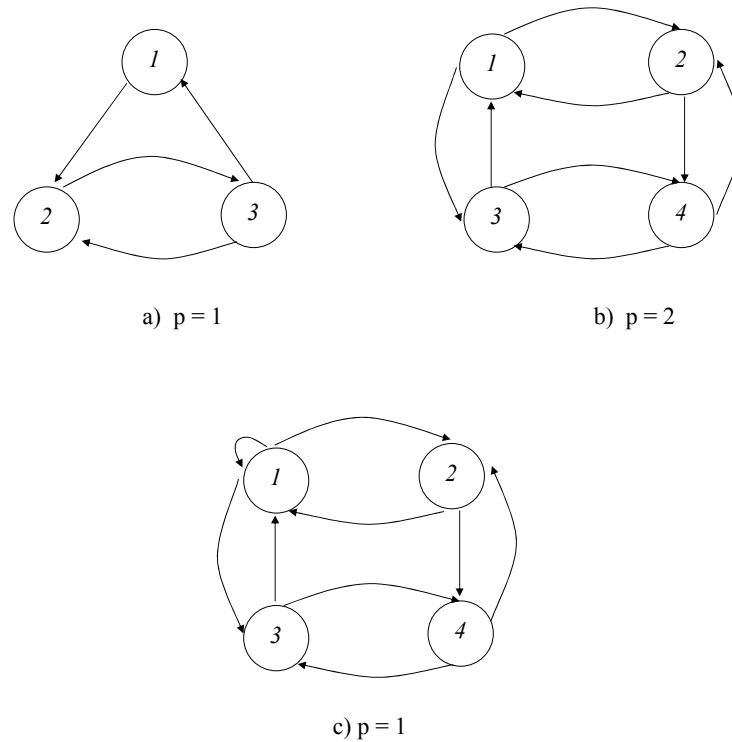


Figura 2.8: Periodicidade das classes finais a), b) e c) (tomado de [15])

- Se a cadeia tem mais de uma classe final e várias classes de passo, trata-se de uma *cadeia não ergódica*

No caso de cadeias de Markov com apenas uma classe final, em regime estacionário P^k terá um período igual ao período da classe final. Assim, só existirá $\lim_{k \rightarrow \infty} P^k$ no caso de classes finais aperiódicas.

Classificação segundo a periodicidade das classes finais

Segundo o número e periodicidade das classes finais tem-se que:

- Se a cadeia tiver uma só classe final aperiódica, será uma cadeia de Markov *regular*. Uma cadeia regular poder ser ergódica ou quase-ergódica.
- Se a cadeia tiver várias classes finais, todas aperiódicas então o $\lim_{k \rightarrow \infty} P^k$ é único. No entanto, as probabilidades em estado estacionário dependem do estado inicial. Essa cadeia será não regular.

- Se a cadeia tiver várias classes finais, todas periódicas então o $\lim_{k \rightarrow \infty} P^k$ não é único. Cadeias deste tipo são denominadas *poli-cíclicas*.
- Se a cadeia tiver várias classes finais e algumas delas periódicas e outras não, será uma cadeia de Markov *mista*.

A seguir apresentam-se alguns exemplos de diagramas de transição para diferentes tipos de cadeias.

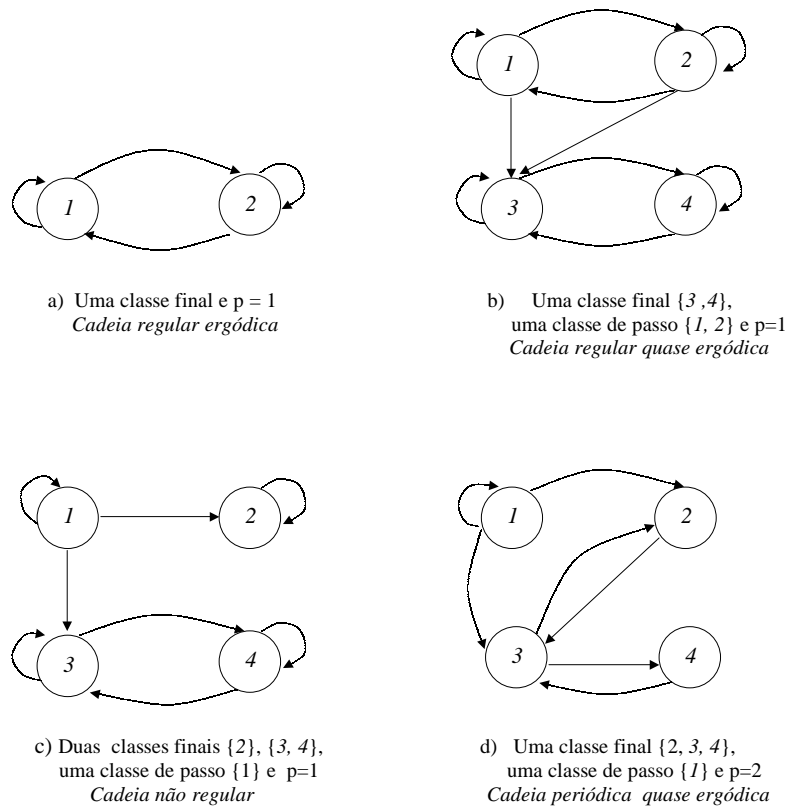


Figura 2.9: Exemplos de classificação de cadeias de Markov (tomado de [15])

Significado da ergodicidade

A ergodicidade tem relação com o comportamento a longo prazo do sistema. Em uma cadeia ergódica, todos os estados têm uma probabilidade não nula de ocorrer em regime estacionário, independente do estado inicial.

Em cadeias quase-ergódicas, dependendo do estado inicial, existem os estados classes de passo que desaparecerão a longo prazo e, portanto só podem aparecer enquanto o sistema não tenha alcançado o regime permanente. O sistema só poderá ocupar os estados das classes de passo, de forma transitória. Uma vez alcançado o regime permanente, o comportamento da cadeia não difere do caso ergódico, uma vez que os estados transitórios foram eliminados.

Em cadeias não ergódicas e não regulares (mais de uma classe final) o comportamento a longo prazo será distinto dependendo da classe final na qual o sistema vá parar. A probabilidade estacionária de cada estado depende do estado inicial.

Significado da periodicidade

Uma classe final aperiódica representa uma situação na qual o sistema evolui entre os distintos estados de forma completamente aleatória, de acordo com as probabilidades indicadas na matriz P . O significado das probabilidades estacionárias representa a probabilidade real de que o sistema encontre-se entre um ou outro estado da classe final.

Portanto, em uma transição qualquer as cadeias regulares e não regulares, representam sistemas que convergem para uma certa probabilidade estacionária, seja única ou não.

Um caso particular importante dentro das classes finais aperiódicas é o denominado *estado absorvente*. Estados absorventes são estados que constituem por si mesmos uma classe final, em que a única transição possível é ir outra vez ao mesmo estado. Na matriz P um estado absorvente é representado por uma linha composta por zeros exceto por um "1" na diagonal principal. O significado do estado absorvente é o de que sistema chegou a uma situação em que não pode mais evoluir.

No caso de classes finais periódicas, pode-se estabelecer uma partição dos seus estados

em p subclasses, de forma que a evolução da cadeia dentro desta classe consistirá na passagem sucessiva por um estado de subclasse 1, um da subclasse 2, etc., até chegar a um estado da subclasse p , depois do qual voltará para um estado da subclasse 1. As classes finais periódicas são um caso intermediário entre um sistema estocástico, em que não se pode prever nada de sua evolução além da mera descrição probabilística, e um sistema determinístico, em que a evolução é conhecida e, portanto pode ser prevista e calculada. Neste caso não se conhece qual será o estado seguinte com exatidão, mas sim que este estado é parte de um subconjunto bem determinado.

2.3.5 Determinação das probabilidades estacionárias

A classificação das cadeias de Markov permite identificar a forma da matriz P em estado estacionário P^* , facilitando-se assim o seu cálculo exato, sem a necessidade de determinar uma potência elevada de P .

De acordo com as definições anteriores, a matriz P^* para uma cadeia ergódica e quase ergódica é uma matriz com todas as linhas iguais, o que facilita seu cálculo. No caso de cadeias não ergódicas, como algumas das linhas de P^* são diferentes das demais, seu cálculo fica mais complexo.

No caso do comportamento desejado do algoritmo LMS, este será projetado de forma a permitir a convergência do coeficiente para o seu valor ótimo \hat{w} , na ausência do ruído de medição, independentemente do valor inicial de $w(n)$. Esse requisito implica que a matriz P_w deverá ter uma única classe final, com um único estado absorvente em $w(n) = \hat{w}$. Assim, P_w deve corresponder a uma cadeia de Markov quase-ergódica [16].

Portanto só entregam-se os passos para determinar P^* desde cadeias ergódicas e quase-ergódicas.

Para todas as cadeias que não tenham classes finais periódicas, considera-se,

$$P^* = \lim_{n \rightarrow \infty} P^n$$

Probabilidades estacionárias para cadeias regulares ergódicas [15]

Como trata-se de uma cadeia regular e ergódica, a probabilidade de que o sistema se encontre em um estado ou em outro depois de um número elevado de transições não depende do estado inicial. Todas as linhas de P^* são iguais. Portanto a matriz é da forma:

$$P^* = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}$$

Sabendo da equação (2.34) que a matriz P^* satisfaz:

$$P^*P = P^*$$

Desenvolvendo esta equação matricial para qualquer uma das linhas da matriz, tem-se que:

$$\sum_{k=1}^n \pi_k p_{ik} = \pi_i \quad \text{para } i = 1, 2, \dots, n \quad (2.35)$$

Além disso, lembrando que para toda matriz estocástica,

$$\sum_{i=1}^n \pi_i = 1 \quad (2.36)$$

O conjunto de equações (2.35) e (2.36) corresponde a um sistema de $n + 1$ equações com n incógnitas, do qual a equação (2.36) nunca deverá ser eliminada, já que esta é a equação que faz P^* estocástica.

Exemplo 2.1.5: Tem-se a seguinte matriz de probabilidades de transições:

$$P_1 = \begin{pmatrix} 0,2 & 0,8 \\ 0,6 & 0,4 \end{pmatrix}$$

e sabe-se que,

$$P^* = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix}$$

Com as equações (2.35) e (2.36),

$$0,2\pi_1 + 0,6\pi_2 = \pi_1$$

$$0,8\pi_1 + 0,4\pi_2 = \pi_2$$

$$\pi_1 + \pi_2 = 1$$

Eliminando uma das primeiras equações, tem-se

$$\pi_1 = 0,429$$

$$\pi_2 = 0,571$$

Probabilidades estacionárias para cadeias regulares quase ergódicas [15]

Para este tipo de cadeias, P^* tem a mesma forma que as matrizes das cadeias regulares ergódicas com a diferença que $\pi_{\text{classes de passo}} = 0$, dado que depois de infinitas transições o sistema estará com certeza em uma classe final.

Exemplo 2.1.6: A matriz de probabilidades de transições P_2 é:

$$P_2 = \begin{pmatrix} 0,5 & 0,4 & 0,1 & 0 \\ 0,3 & 0,3 & 0,4 & 0 \\ 0 & 0 & 0,2 & 0,8 \\ 0 & 0 & 0,6 & 0,4 \end{pmatrix}$$

e a matriz de probabilidades estacionárias é,

$$P_1 = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix}$$

Considerando as equações (2.35) e (2.36) para determinar π_1, π_2, π_3 e π_4

$$0,5\pi_1 + 0,3\pi_2 = \pi_1$$

$$0,4\pi_1 + 0,3\pi_2 = \pi_2$$

$$0,1\pi_1 + 0,4\pi_2 + 0,2\pi_3 + 0,6\pi_4 = \pi_3$$

$$0,8\pi_3 + 0,4\pi_4 = \pi_4$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

Eliminando uma das primeiras quatro equações têm-se,

$$\pi_1 = 0$$

$$\pi_2 = 0$$

$$\pi_3 = 0,429$$

$$\pi_4 = 0,571$$

Para este exemplo, o diagrama de probabilidades de transição corresponde ao da Figura

(2.9b), em que os estados $\{1, 2\}$ formam uma classe de passo. Assim, a longo prazo o processo não se encontrará em nenhum destes estados, já que $\pi_1 = \pi_2 = 0$. Obviamente, estes estados de passo podem ser eliminados diretamente das equações, caso sejam conhecidos antecipadamente. Neste caso, só duas incógnitas precisariam ser determinadas, simplificando o cálculo. Aqui, os estados de classe final são $\{3, 4\}$. Portanto para determinar π_3 e π_4 basta considerar a parte de P_2 que corresponde a estes estados, ou seja

$$\hat{P}_2 = \begin{pmatrix} 0,2 & 0,8 \\ 0,6 & 0,4 \end{pmatrix}$$

Observe que \hat{P}_2 é igual à matriz P_1 do *Exemplo 2.1.5*, onde as probabilidades estacionárias resultaram $\pi_1 = 0,429$ e $\pi_2 = 0,571$, que são as mesmas para o *Exemplo 2.1.6*

2.4 CONCLUSÕES

Neste capítulo apresentou-se uma breve revisão do conceito de variável aleatória e cadeias de Markov a serem utilizados no próximo capítulo.

Conclui-se que para o comportamento desejado do algoritmo LMS com entradas e coeficientes com precisão finita, os coeficientes do filtro deverão ser modelados como uma cadeia de Markov quase-ergódica.

No próximo capítulo desenvolve-se a função densidade de probabilidade (fdp) do coeficiente adaptativo, condicionada ao coeficiente anterior e ao estado do sinal de entrada. A partir da fdp dos coeficientes determina-se a matriz de probabilidades de transição da cadeia de Markov. Exemplos de comparação entre os desempenhos obtidos pelo modelo e pela simulação são apresentados.

Capítulo 3

ALGORITMO LMS EM PRECISÃO FINITA

3.1 INTRODUÇÃO

Sendo o algoritmo adaptativo LMS, por sua simplicidade e robustez, um dos mais populares na implementação digital de sistemas adaptativos, é importante estudar o seu comportamento quando implementado em precisão finita. Vários estudos têm sido feitos (ver Capítulo 2), mas uma análise exata do comportamento do algoritmo tem sido complicada, devido a dificuldades matemáticas na análise estatística.

Neste capítulo determinam-se as estatísticas de $w(n)$ considerando sinais e coeficientes representados em precisão finita e incluindo os efeitos de saturação. O estudo é feito para o caso de comprimento 1 (um único coeficiente $w(n)$) para evitar a complexidade algébrica inerente ao caso multidimensional. O estudo deste caso, no entanto, permite a obtenção de informações importantes sobre o comportamento do algoritmo, que não poderiam ser obtidas usando os modelos existentes na literatura. O coeficiente adaptativo é modelado como uma cadeia de Markov. A sua função densidade de probabilidade (fdp) é determinada, condicionada ao valor do coeficiente anterior e ao estado do sinal de entrada. A partir da fdp dos coeficientes, determina-se a matriz de probabilidades de transição da cadeia de Markov para, no próximo capítulo, estudar através de suas propriedades, o comportamento do algoritmo em regime permanente. Valida-se o modelo proposto, apresentando exemplos de comparação

entre os desempenhos previstos pelo modelo teórico e obtidos por simulação.

3.2 CONSIDERAÇÕES INICIAIS

Considere o diagrama em blocos mostrado na Figura 3.1, referente ao problema de identificação de sistemas.

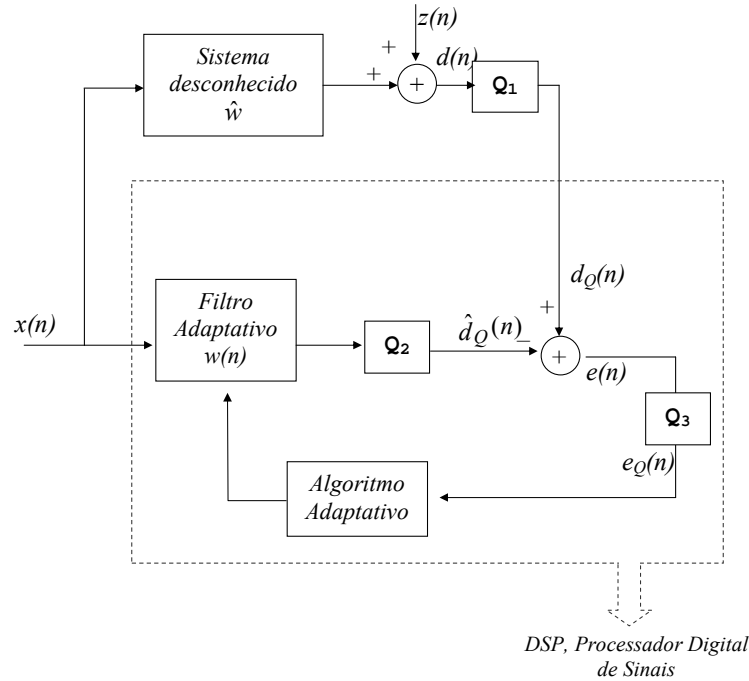


Figura 3.1: Identificação de sistemas implementada em precisão finita

Na Figura 3.1, $x(n)$ é o sinal de entrada, $z(n)$ é um ruído aditivo, considerado branco, gaussiano e de média zero e $d(n)$ é o sinal desejado. \hat{w} é o coeficiente do sistema desconhecido e $w(n)$ é o coeficiente do filtro adaptativo. Assume-se que $x(n)$ é um sinal aleatório discreto com $-1 \leq x(n) \leq (1 - \Delta)$, quantizado com b bits, e com todos os níveis equiprováveis. Este modelo é compatível com os utilizados em transmissão digital [17].

Assume-se também que os coeficientes do filtro adaptativo e as operações internas do sistema (dentro da caixa tracejada) são realizadas com b bits de precisão. Os blocos Q_1 , Q_2 e Q_3 são quantizadores de b bits com passo uniforme $\Delta = 2^{1-b}$ e com saídas limitadas em $[-1, 1 - \Delta]$. A Fig. 3.2 mostra a função de transferência do quantizador para $b = 3$. O sinal

$d_Q(n)$ é a versão quantizada de $d(n)$, $\hat{d}_Q(n)$ é a saída do filtro adaptativo quantizada em b bits¹ e $e_Q(n)$ é o erro de estimação representado em b bits.

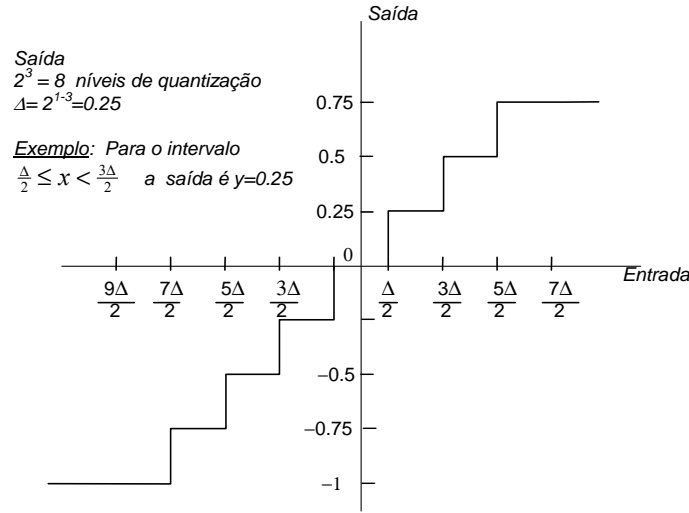


Figura 3.2: Relação entrada e saída para um quantizador de 3 bits

Matematicamente, a fdp de $x(n)$ é dada por

$$f_x(x) = \frac{1}{N} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \delta(x - k\Delta) \quad (3.1)$$

em que $N = 2^b$. A fdp de $z(n)$ é dada por

$$f_z(z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\left(\frac{z}{\sqrt{2}\sigma_z}\right)^2} \quad (3.2)$$

O coeficiente $w(n)$ do filtro adaptativo é um sinal escalar discreto e distribuído em $[-1, 1 - \Delta]$.

¹ Em geral, a saída terá pelo menos $2b$ bits antes de \mathbf{Q}_2 , devido ao maior comprimento de palavra do acumulador em processadores digitais.

3.2.1 Equações do algoritmo LMS em precisão finita

Considerando a Fig. 3.1 e a estrutura do algoritmo LMS [1], as equações que definem a atualização dos coeficientes do filtro adaptativo em precisão finita são:

$$\begin{aligned}
 d_Q(n) &= Q\{\hat{w}x(n) + z(n)\} \\
 \hat{d}_Q(n) &= Q\{w(n)x(n)\} \\
 e(n) &= d_Q(n) - \hat{d}_Q(n) \\
 e_Q(n) &= Q\{d_Q(n) - \hat{d}_Q(n)\} \\
 y_Q(n) &= Q\{\mu e_Q(n)x(n)\} \\
 w(n+1) &= Q\{w(n) + y_Q(n)\}
 \end{aligned} \tag{3.3}$$

Note que a definição de $e_Q(n)$ e a inclusão da operação de quantização na definição de $w(n+1)$ tornam-se necessárias devido à possibilidade de saturação, uma vez que as amplitudes de $e(n)$ e de $w(n) + y_Q(n)$ podem extrapolar os limites de saturação do quantizador.

Para estudar a evolução dinâmica do coeficiente $w(n)$ do filtro adaptativo, determina-se a fdp de $w(n+1)$ condicionada em $w(n)$. Para isso, deve-se determinar diversas fdp relativas aos diferentes termos em (3.3). Esse trabalho é apresentado na próxima seção.

3.3 DETERMINAÇÃO DA ESTATÍSTICA DE $w(n)$

Da equação (3.3) verifica-se que a fdp de $w(n+1)$ depende da estatística de $y_Q(n)$, a qual depende da estatística de $e_Q(n)$ que, por sua vez, depende das estatísticas de $d_Q(n)$ e de $\hat{d}_Q(n)$. Inicia-se o estudo pela fdp de $d_Q(n)$.

3.3.1 Determinação da fdp de $d_Q(n)$

De (3.3),

$$\begin{aligned} d_Q(n) &= Q\{d(n)\} \text{ com} \\ d(n) &= \hat{w}x(n) + z(n) \end{aligned}$$

em que $d(n)$ é o sinal desejado não quantizado.

Dado \hat{w} , define-se o sinal de saída do filtro ótimo como $x_o(n) = \hat{w}x(n)$, sendo que

$Pr\{\mathbf{x}_o = x_o\} = Pr\{\hat{w}x = x_o\} = Pr\{x = \frac{x_o}{\hat{w}}\}$. Assim, tem-se

$$f_{\mathbf{x}_o}(x_o) = \frac{1}{N} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \delta(x_o - \hat{w}k\Delta) \quad (3.4)$$

Como $d(n)$ é a soma de duas variáveis aleatórias independentes, sua fdp será dada pela convolução das fdp individuais (Capítulo 2, Seção 2.2.5). Assim, como $z(n) = d(n) - x_o(n)$,

$$f_d(d) = \int_{-\infty}^{\infty} f_{\mathbf{x}_o}(x_o) f_z(d - x_o) dx_o \quad (3.5)$$

Substituindo (3.2) e (3.4) em (3.5), e com $x_o = \hat{w}k\Delta$, a função densidade de probabilidade de $d(n)$ não quantizado, $f_d(d)$ fica:

$$f_d(d) = \frac{1}{N\sigma_z\sqrt{2\pi}} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} e^{-\left(\frac{d-\hat{w}k\Delta}{\sqrt{2}\sigma_z}\right)^2} \quad (3.6)$$

Para determinar a fdp de $d_Q(n)$, deve-se determinar a probabilidade de que $d(n)$ esteja no i -ésimo intervalo de quantização $[d_{1_i}, d_{2_i}]$. Denominando essa probabilidade D_i , tem-se

$$D_i = \int_{d_{1_i}}^{d_{2_i}} f_d(d) dd \quad (3.7)$$

Substituindo (3.6) em (3.7) tem-se que

$$D_i = \frac{1}{N_x \sigma_z \sqrt{2\pi}} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \int_{d_{1_i}}^{d_{2_i}} e^{-\left(\frac{d-\hat{w}k\Delta}{\sqrt{2}\sigma_z}\right)^2} dd \quad (3.8)$$

reordenando (3.8)

$$D_i = \frac{1}{N} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \frac{1}{\sigma_z \sqrt{2\pi}} \int_{d_{1_i}}^{d_{2_i}} e^{-\left(\frac{d-\hat{w}k\Delta}{\sqrt{2}\sigma_z}\right)^2} dd$$

em que,

$$\frac{1}{\sigma_z \sqrt{2\pi}} \int_{d_{1_i}}^{d_{2_i}} e^{-\left(\frac{d-\hat{w}k\Delta}{\sqrt{2}\sigma_z}\right)^2} dd \quad (3.9)$$

é uma função conhecida e desenvolvida no capítulo 2, seção 2.2.3. Assim, tem-se,

$$\frac{1}{\sigma_z \sqrt{2\pi}} \int_{d_{1_i}}^{d_{2_i}} e^{-\left(\frac{d-\hat{w}k\Delta}{\sqrt{2}\sigma_z}\right)^2} dd = \text{erf}\left(\frac{d_{2_i} - k\hat{w}\Delta}{\sigma_z}\right) - \text{erf}\left(\frac{d_{1_i} - k\hat{w}\Delta}{\sigma_z}\right) \quad (3.10)$$

então,

$$D_i = \frac{1}{N} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \left[\text{erf}\left(\frac{d_{2_i} - k\hat{w}\Delta}{\sigma_z}\right) - \text{erf}\left(\frac{d_{1_i} - k\hat{w}\Delta}{\sigma_z}\right) \right] \quad (3.11)$$

e a fdp é dada por:

$$f_{\mathbf{d}_Q}(d_Q) = \sum_{i=-2^{(b-1)}}^{i=2^{(b-1)}-1} D_i \delta(d_Q - i\Delta_x) \quad (3.12)$$

com $d_{2_i} = i\Delta + 0,5\Delta$ e $d_{1_i} = i\Delta - 0,5\Delta$ (note que o limite inferior do primeiro intervalo é $-\infty$ e, analogamente, o limite superior do último intervalo é $+\infty$).

Conhecida a fdp de d_Q , é possível determinar a fdp do erro quantizado.

3.3.2 Determinação da fdp condicional de $e_Q(n)$

Da Figura 3.1 e de (3.3),

$$e(n) = d_Q(n) - \hat{d}_Q(n) \quad (3.13)$$

em que,

$$\hat{d}_Q(n) = Q\{w(n)x(n)\}$$

As variáveis aleatórias $d_Q(n)$ e $Q\{w(n)x(n)\}$ dependem da variável aleatória $x(n)$ e do coeficiente $w(n)$ do filtro. Assim, a fdp de interesse é $f_{e_Q}(e_Q|k\Delta, w(n))$.

Note que o resultado de $e(n)$ em (3.13) pode superar o máximo ou o mínimo valores estabelecidos pela reposta do quantizador. Assim, para $e(n) < -1$ os valores de $e_Q(n)$ ficarão saturados nos valores extremos da reposta do quantizador, o mesmo acontece para $e_Q(n) > (1 - \Delta)$.

Inicialmente, deve-se determinar a fdp de $e(n)$. Considerando que,

$$\begin{aligned} d_{Q\max} &= (2^{(b-1)} - 1)\Delta \\ Q\{w(n)x(n)\}_{\min} &= -2^{(b-1)}\Delta = -1 \end{aligned}$$

tem-se que,

$$\begin{aligned} \{e(n)\}_{\max} &= (2^b - 1)\Delta \text{ e} \\ \{e(n)\}_{\min} &= -(2^b - 1)\Delta \end{aligned}$$

A fdp de $e(n)$ será dada por:

$$f_e\{e|k\Delta, w(n)\} = \sum_{k_e=-2^b+1}^{k_e=2^b-1} E_{k_e} \delta(e - k_e\Delta)$$

em que E_{k_e} representa a probabilidade de ocorrência do k_e -ésimo estado de $e(n)$. A seguir

apresentam-se os passos para determinar E_{k_e} . Seja

$$\hat{d}_1(n) = \text{round}\{w(n)k\}\Delta$$

em que $\text{round}\{x\}$, para $m \leq x < m + 1$, $m \in \mathbb{Z}$, é:

$$\lfloor x \rfloor = \begin{cases} m & \text{se } x < m + \frac{1}{2} \\ m + 1 & \text{se } x \geq m + \frac{1}{2} \end{cases}$$

Como $\hat{d}_1(n)$ pode exceder os limites do quantizador quando $w(n) = -1$ e $x(n) = -1$, então seja:

$$\hat{d}(n) = \hat{d}_1 - \Delta\delta(1 - \hat{d}_1)$$

Então, $e(n) = d_Q(n) - \hat{d}(n)$ e

$$\begin{aligned} \Pr\{\mathbf{e} = e(n) | x(n) = k\Delta, w(n)\} &= \Pr\{d_Q - \hat{d}(n) = e | k\Delta, w(n)\} \\ &= \Pr\{d_Q = e + \hat{d}(n) | k\Delta, w(n)\} \end{aligned}$$

Usando $\Pr\{d_Q\}$ determinada a partir de (3.12), tem-se:

$$E_{k_e} = \frac{1}{N} \left[\text{erf} \left(\frac{d_{2k_e} - \hat{w}x}{\sigma_z} \right) - \text{erf} \left(\frac{d_{1k_e} - \hat{w}x}{\sigma_z} \right) \right]$$

com (novamente, excetuando o limite inferior do primeiro intervalo e o limite superior do último intervalo)

$$\begin{aligned} d_{2k_e} &= d_Q(n) + 0,5\Delta = k_e\Delta + \hat{d}(n) + 0,5\Delta, \quad \text{e} \\ d_{1k_e} &= d_Q(n) - 0,5\Delta = k_e\Delta + \hat{d}(n) - 0,5\Delta. \end{aligned}$$

Como $e_Q(n)$ é obtido de $e(n)$ pela saturação dos valores que excedem os limites do quan-

tizador, as probabilidades de ocorrência destes valores devem somar-se às probabilidades de ocorrência dos valores extremos na distribuição de $e_Q(n)$. Assim, a fdp de $e_Q(n)$ é dada por:

$$f_{e_Q}(e_Q|k\Delta, w(n)) = \sum_{k_1=-2^{(b-1)}-1}^{k_1=2^{(b-1)}-1} \Pr\{e_Q = k_1\Delta|k\Delta, w(n)\}\delta(e_Q - k_1\Delta) \quad (3.14)$$

em que

$$\Pr\{e_Q = k_1\Delta|k\Delta, w(n)\} = \begin{cases} \Pr\{e = k_1\Delta|k\Delta, w(n)\}, & -1 < e_Q(n) < 1 - \Delta \\ \sum_{\substack{k_e=2^b-1 \\ k_e=k_1}} \Pr\{e = k_e\Delta|k\Delta, w(n)\}, & e_Q(n) = 1 - \Delta \\ \sum_{k_e=-2^b+1} \Pr\{e = k_e\Delta|k\Delta, w(n)\}, & e_Q(n) = -1 \end{cases} \quad (3.15)$$

3.3.3 Determinação da fdp de $y_Q(n)$

Considere $y(n) = \mu e_Q(n)x(n)$, sendo que $e_Q(n) = k_e\Delta$ e $x(n) = k\Delta$ são valores quantizados. Assim,

$$\begin{aligned} y &= \mu k_e \Delta k \Delta \\ y &= \mu (k_e k) \Delta^2 \\ y &= \mu k_y \Delta_y \end{aligned} \quad (3.16)$$

em que $k_y = k_e k$ e $\Delta_y = \Delta^2$

Como o $\max\{k_y\}$ acontece para $\min\{k_e\}$ e $\min\{k\}$ então,

$$k_{y_{max}} = 2^{2(b-1)}$$

Agora o $\min\{k_y\}$ acontece para $\min\{k_e\}$ e $\max\{k\}$ ou vice versa, assim

$\min\{k_y\} = -2^{(b-1)}(2^{b-1} - 1)$, ou bem

$$k_{y_{min}} = -2^{(b-2)}(2^b - 2)$$

Sendo $y(n) = \mu e_Q(n)x(n)$ tem-se que,

$$e_Q(n) = \frac{y(n)}{\mu x(n)} = \frac{\mu k_y \Delta^2}{\mu x(n)} = \frac{\mu k_y \Delta^2}{\mu k \Delta} = \frac{k_y \Delta}{k}$$

Assim, as probabilidades para $y(n)$ não quantizado e condicionado em $k\Delta$ e $w(n)$ são dadas por:

$$\begin{aligned} Pr\{\mathbf{y}(\mathbf{n}) = y(n) | x(n) = k\Delta, w(n)\} &= Pr\{\mu e_Q(n)x(n) = y | k\Delta, w(n)\} \\ &= Pr\left\{e_Q(n) = \frac{y(n)}{\mu x(n)} | k\Delta, w(n)\right\} \\ &= Pr\left\{e_Q(n) = \frac{k_y \Delta}{k} | k\Delta, w(n)\right\} \end{aligned}$$

em que $Pr\{e_Q | k\Delta, w(n)\}$ pode ser determinada das equações (3.14) e (3.15). Assim a fdp de $y(n)$ fica:

$$f_y(y | k\Delta_x, w(n)) = \sum_{k_{y_{min}}}^{k_{y_{max}}} Pr\left\{e_Q = \frac{k_y \Delta}{k} | k\Delta_x, w(n)\right\} \delta(y - k_y \Delta_y)$$

Como $y(n)$ é o resultado de uma multiplicação de números de b bits seu valor será representado com $2b$ bits. $y_Q(n)$ será o valor de $y(n)$ quantizado com b bits. Assim, a fdp de y_{Q_n} será:

$$f_{y_Q}\{y_Q | k\Delta, w(n)\} = \sum_{k_{y_Q} = -2^{(b-1)}}^{k_{y_Q} = 2^{(b-1)} - 1} Pr\{y_Q = k_{y_Q} \Delta | k\Delta, w(n)\} \delta(y_Q - k_{y_Q} \Delta_x) \quad (3.17)$$

com

$$\begin{aligned} & \Pr\{y_Q = k_{y_Q}\Delta | k\Delta, w_n\} \\ &= \Pr\{y \geq y_1 | k\Delta\}u(y - y_1) - \Pr\{y \geq y_2 | k\Delta\}u(y - y_2) \end{aligned} \quad (3.18)$$

para,

$$\begin{aligned} y_1 &= k_{y_Q}\Delta - 0,5\Delta \\ y_2 &= k_{y_Q}\Delta + 0,5\Delta \end{aligned}$$

e $u(\cdot)$ representando a função degrau unitário. Conhecida a fdp de $y_Q(n)$ podemos determinar a estatística de $w(n+1)$.

3.3.4 Determinação da fdp de $w(n+1)$

Para simplificar as equações, definimos a variável $\omega(n+1) = w(n) + y_Q(n)$. Assim,

$$\omega(n+1) = w(n) + Q\{\mu e_Q(n)x(n)\} = w(n) + Q\{y(n)\} \quad (3.19)$$

e

$$\begin{aligned} \Pr\{\omega(n+1) = k_{w1}\Delta | x(n) = k\Delta, w(n)\} &= \Pr\{w(n) + y_Q = k_{w1}\Delta | k\Delta, w(n)\} \\ &= \Pr\{y_Q = k_{w1}\Delta - w(n) | k\Delta, w(n)\} \end{aligned} \quad (3.20)$$

Essa probabilidade pode ser determinada a partir de (3.17) e (3.18). No entanto, para determinarmos a estatística de $w(n+1)$ precisamos incluir os efeitos de saturação em (3.20).

Logo,

$$\begin{aligned}
 & f_{\mathbf{w}} \{w(n+1)|k\Delta, w(n)\} \\
 &= \sum_{k_w=-2^{b-1}}^{k_w=2^{b-1}-1} \Pr\{w(n+1) = k_w\Delta|k\Delta, w(n)\} \delta(w(n+1) - k_w\Delta)
 \end{aligned} \tag{3.21}$$

em que

$$\begin{aligned}
 & \Pr\{w(n+1) = k_w\Delta|k\Delta, w(n)\} \\
 &= \begin{cases} \Pr\{\omega(n+1) = k_w\Delta|k\Delta, w(n)\}, & -1 < w(n+1) < 1 - \Delta \\ \sum_{\substack{k_{w1}=2^b-2 \\ k_{w1}=k_w}} \Pr\{\omega(n+1) = k_{w1}\Delta|k\Delta, w(n)\}, & w(n+1) = 1 - \Delta \\ \sum_{k_{w1}=-2^b} \Pr\{\omega(n+1) = k_{w1}\Delta|k\Delta, w(n)\}, & w(n+1) = -1 \end{cases}
 \end{aligned} \tag{3.22}$$

Essa equação permite a determinação dos elementos da matriz de transição P_w da cadeia de Markov. A seguir, apresenta-se um resumo das equações do modelo.

3.4 RESUMO: MODELO PROPOSTO PARA

$$f_{\mathbf{w}} \{w(n+1)|x(n) = k\Delta, w(n)\}$$

$$\begin{aligned}
 f_{\mathbf{w}} \{w(n+1)|k\Delta, w(n)\} &= \sum_{k_w=-2^{(b-1)}}^{k_w=2^{(b-1)}-1} Pr\{w(n+1) = k_w\Delta|k\Delta, w(n)\} \delta(w(n+1) - k_w\Delta)
 \end{aligned} \tag{3.23}$$

com,

$$\begin{aligned} & \Pr\{w(n+1) = k_w \Delta | k\Delta, w_n\} \\ &= \begin{cases} \Pr\{\omega(n+1) = k_w \Delta | k\Delta, w(n)\}, & -1 < w(n+1) < 1 - \Delta \\ \sum_{\substack{k_{w1}=2^b-2 \\ k_{w1}=k_w}} \Pr\{\omega(n+1) = k_{w1} \Delta | k\Delta, w(n)\}, & w(n+1) = 1 - \Delta \\ \sum_{k_{w1}=-2^b} \Pr\{\omega(n+1) = k_{w1} \Delta | k\Delta, w(n)\}, & w(n+1) = -1 \end{cases} \end{aligned} \quad (3.24)$$

sendo,

$$\begin{aligned} \Pr\{\omega(n+1) = k_w \Delta | k\Delta, w(n)\} &= \Pr\{y_Q(n) + w(n) = \omega(n+1) | k\Delta, w(n)\} \\ &= \Pr\{y_Q(n) = \omega(n+1) - w(n) | k\Delta, w(n)\} \end{aligned}$$

como,

$$f_{y_Q} \{y_Q | k\Delta, w(n)\} = \sum_{k_{yQ}=-2^{(b-1)}}^{k_{yQ}=2^{(b-1)}-1} \Pr\{y_Q = k_{yQ} \Delta | k\Delta, w(n)\} \delta(y_Q - k_{yQ} \Delta_x) \quad (3.25)$$

e,

$$\begin{aligned} & \Pr\{y_Q = k_{yQ} \Delta | k\Delta, w_n\} \\ &= \Pr\{y \geq y_1 | k\Delta\} u(y - y_1) - \Pr\{y \geq y_2 | k\Delta\} u(y - y_2) \end{aligned} \quad (3.26)$$

para,

$$y_1 = k_{yQ} \Delta - 0,5\Delta$$

$$y_2 = k_{yQ} \Delta + 0,5\Delta$$

Agora, como

$$\begin{aligned}\Pr\{\mathbf{y}(\mathbf{n}) = y(n)|k\Delta, w(n)\} &= \Pr\{\mu e_Q(n)x(n) = y(n)|k\Delta, w(n)\} \\ &= \Pr\{e_Q(n) = \frac{y(n)}{\mu x(n)}|k\Delta, w(n)\}\end{aligned}$$

com,

$$f_{\mathbf{e}_Q}(e_Q|x, w) = \sum_{k_1=-2^{(b-1)}}^{k_1=2^{(b-1)}-1} \Pr\{e_Q = k_1\Delta|x, w\}\delta(e_Q - k_1\Delta_x) \quad (3.27)$$

em que,

$$\begin{aligned}\Pr\{e_Q = k_1\Delta|k\Delta, w(n)\} \\ = \begin{cases} \Pr\{e = k_1\Delta|k\Delta, w(n)\}, & -1 < e_Q(n) < 1 - \Delta \\ \sum_{\substack{k_e=2^b-1 \\ k_e=k_1}} \Pr\{e = k_e\Delta|k\Delta, w(n)\}, & e_Q(n) = 1 - \Delta \\ \sum_{k_e=-2^b+1} \Pr\{e = k_e\Delta|k\Delta, w(n)\}, & e_Q(n) = -1 \end{cases} \quad (3.28)\end{aligned}$$

com,

$$\begin{aligned}\Pr\{e = k_e\Delta|k\Delta, w(n)\} &= \Pr\{d_Q(n) - Q\{w(n)x(n)\} = e|k\Delta, w(n)\} \\ &= \Pr\{d_Q(n) = e1(n) - Q\{w(n)x(n)\}|k\Delta, w(n)\}\end{aligned}$$

sendo que,

$$f_{\mathbf{d}_Q}(d_Q) = \sum_{ki=-2^{(b-1)}}^{ki=2^{(b-1)}-1} D_i\delta(d_Q - ki\Delta_x) \quad (3.29)$$

e

$$D_i = \frac{1}{N} \sum_{k=-2^{(b-1)}}^{k=2^{(b-1)}-1} \left[\operatorname{erf} \left(\frac{d_{2i} - k\hat{w}\Delta}{\sigma_z} \right) - \operatorname{erf} \left(\frac{d_{1i} - k\hat{w}\Delta}{\sigma_z} \right) \right] \quad (3.30)$$

3.5 VERIFICAÇÃO DO MODELO PROPOSTO

Esta seção apresenta resultados de simulação para verificação dos resultados teóricos obtidos. Apresentam-se histogramas comparativos para algumas das equações descritas no modelo derivado.

3.5.1 Exemplos para $f_d(d)$

A seguir apresentam-se alguns gráficos da fdp do sinal desejado $d(n)$, obtido com o modelo teórico dado por (3.6) e os resultados obtidos por simulação. O sinal desejado é definido pela expressão $d(n) = \hat{w}x + z$, em que $\hat{w} = 0,25$, $x(n)$ é o sinal de entrada quantizado em b bits e $z(n)$ um ruído aditivo gaussiano de variância σ_z^2 . Para as simulações considera-se a média de 300000 realizações. As figuras 3.3, 3.4 e 3.5 mostram os resultados das simulações Monte Carlo e o comportamento de $d(n)$ previsto pela eq.(3.6). Considera-se, nas duas primeiras figuras, $b = 2$ bits para a quantização de $x(n)$, ou seja, $x = \{-1; -0,5; 0; 0,5\}$ níveis, e $\sigma_z = 10^{-1}$ e 10^{-3} . Na medida que σ_z é menor é possível observar a quantidade de estados de $x(n)$: no caso da Figura 3.4 quatro estados, e no caso da Figura 3.5 para $b = 3$ bits, oito estados. Pode-se observar nas três figuras que existe muito boa combinação entre os resultados da simulação e a previsão do modelo teórico.

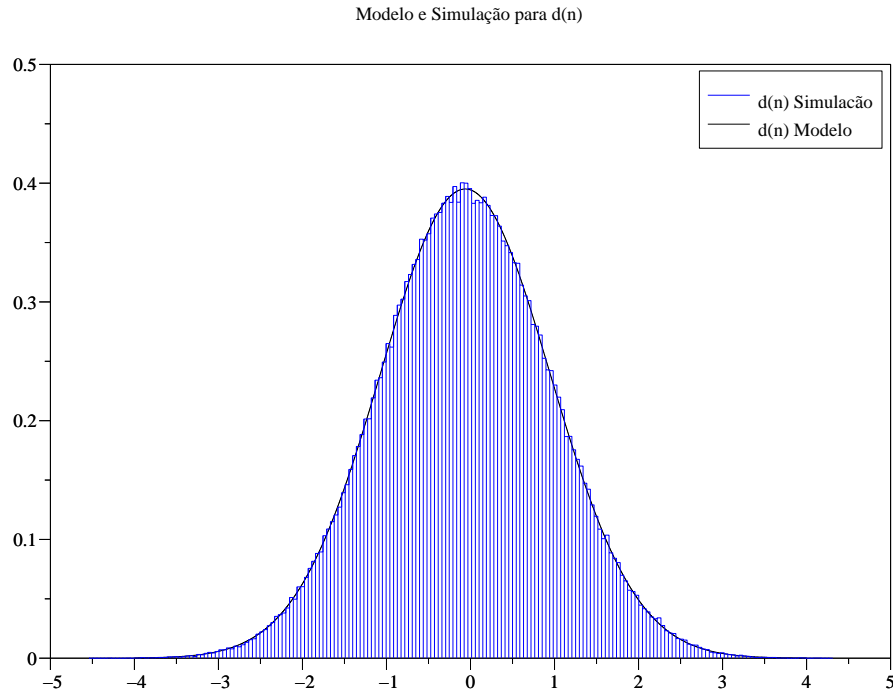


Figura 3.3: Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 2bits$, $\sigma_z = 10^{-1}$ e $\hat{w} = 0,25$

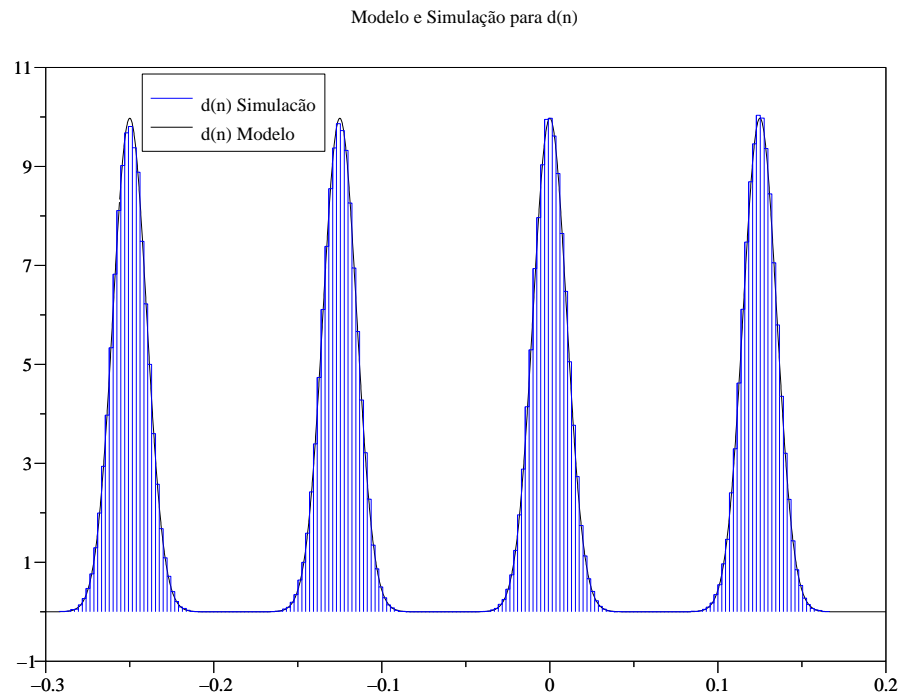


Figura 3.4: Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 2bits$, $\sigma_z = 10^{-3}$ e $\hat{w} = 0,25$

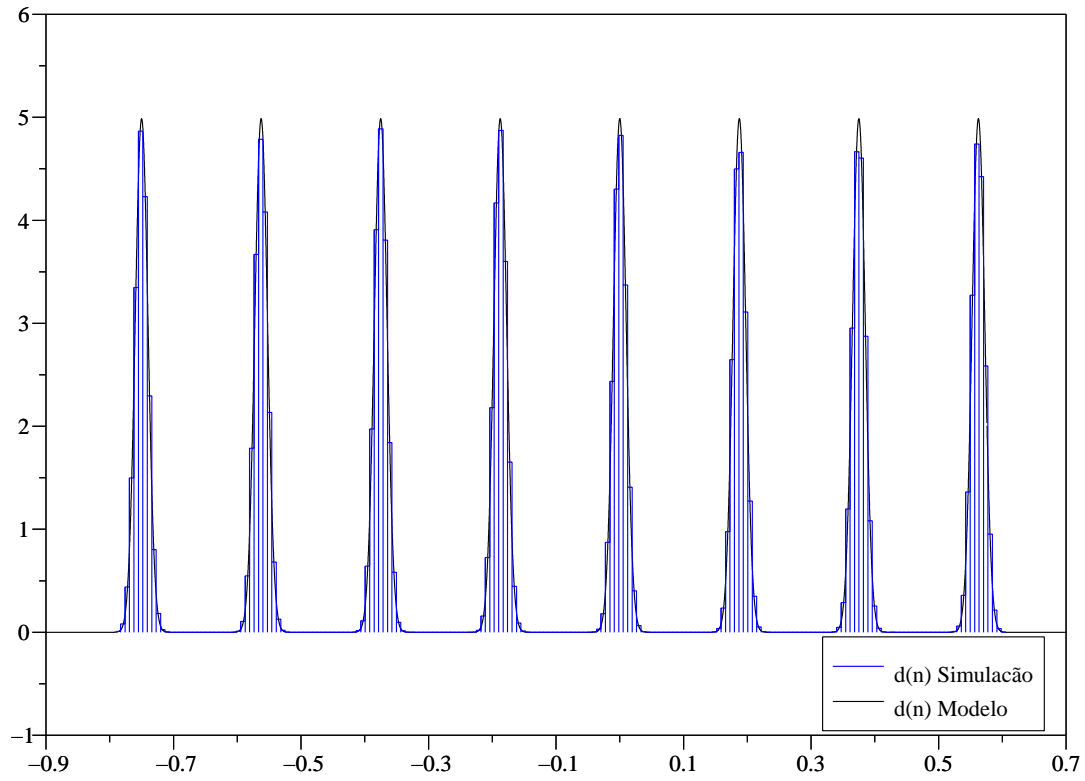
Modelo e Simulação para $d(n)$ 

Figura 3.5: Comparação entre o Modelo e a Simulação para $f_d(d)$, com $b = 3bits$, $\sigma_z = 10^{-3}$ e $\hat{w} = 0,75$

3.5.2 Exemplos para $f_{e_Q}(e_Q)$

A seguir apresentam-se histogramas que comparam as probabilidades de ocorrência de $e_Q(n)$ obtidas pelas equações (3.14) e (3.15) com os resultados obtidos por simulação. O erro quantizado é definido por $e_Q(n) = Q\{d_Q(n) - Q\{w(n)x(n)\}\}$, em que o sinal desejado quantizado é $d_Q(n) = Q\{\hat{w}x(n) + z(n)\}$ e $z(n)$ é um ruído aditivo gaussiano com $\sigma_z = 10^{-4}$. As Figuras 3.6, 3.7, 3.8 e 3.9 mostram os resultados das simulações de Monte Carlo e o comportamento de $e_Q(n)$ previsto pelas eqs.(3.14) e (3.15). Considera-se para a simulação a média de 100000 realizações, para diferentes valores de \hat{w} e condição inicial $w(0)$. Observe nas Figuras 3.6, 3.7 e 3.9 em que $e_Q(n)$ é quantizado em $b = 4$ bits, a presença dos 16 níveis e a probabilidade de ocorrência para cada nível. Em todos os gráficos pode-se observar a grande semelhança entre os resultados da simulação e a previsão do modelo teórico.

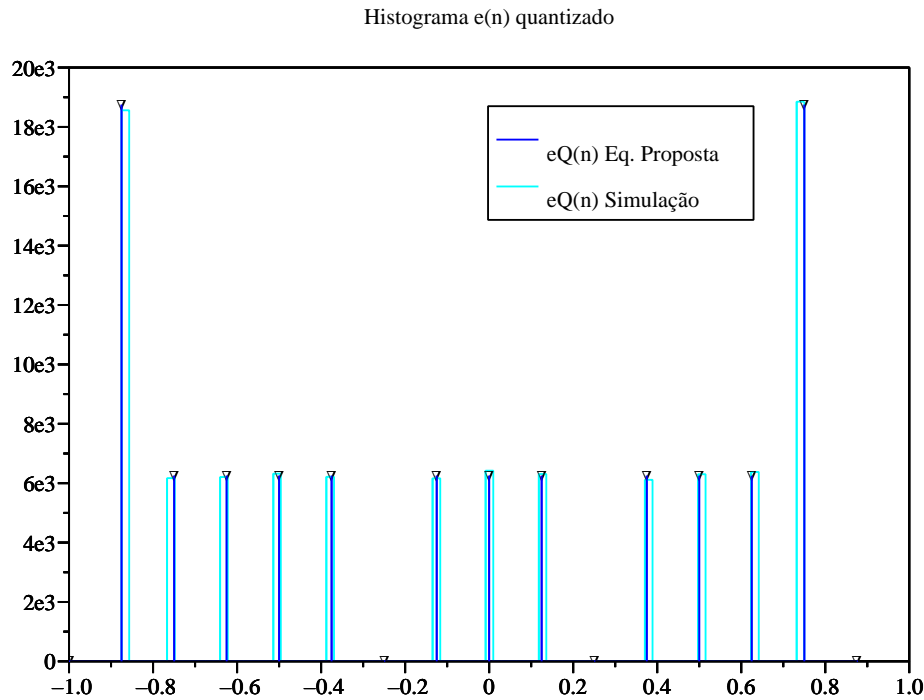


Figura 3.6: Comparação entre o Modelo e a Simulação para $f_{e_Q}(e_Q)$ com $b = 4\text{bits}$, $\sigma_z = 10^{-4}$, $\hat{w} = 1,425$ e $w(0) = 0,125$

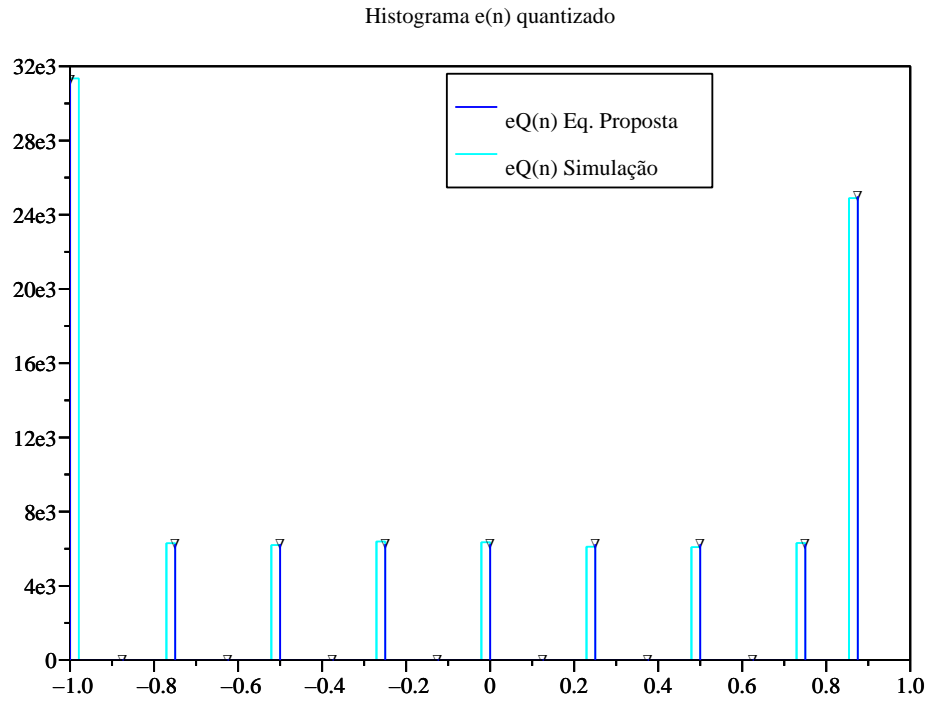


Figura 3.7: Comparação entre o Modelo e a Simulação para $f_{e_Q}(e_Q)$ com $b = 4bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 1$ e $w(0) = -1$

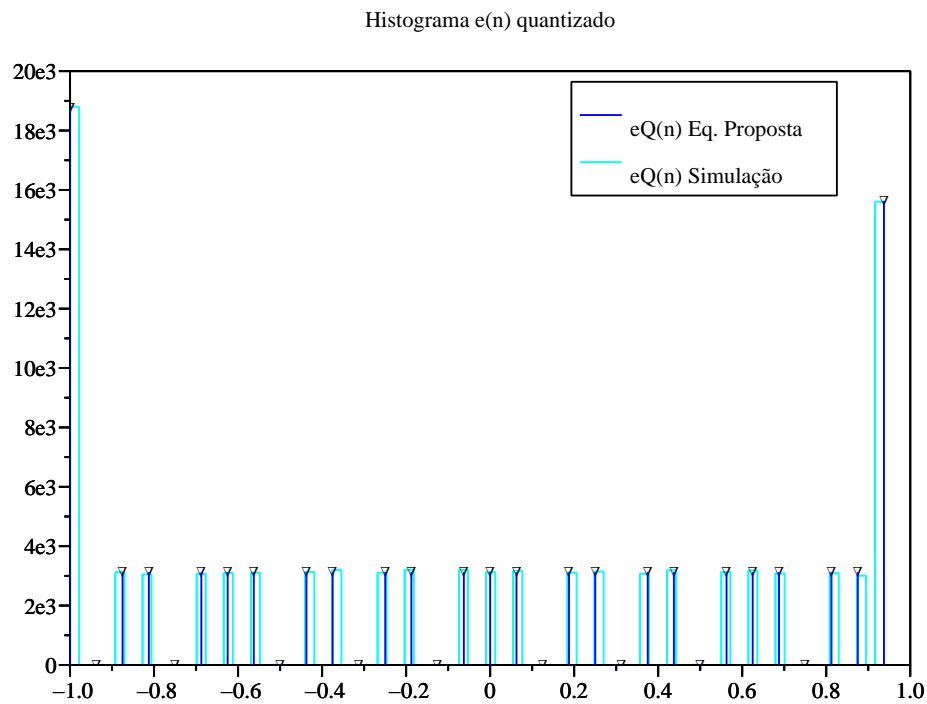


Figura 3.8: Comparação entre o Modelo e a Simulação para $f_{e_Q}(e_Q)$ com $b = 5bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 1,425$ e $w(0) = -0,03125$

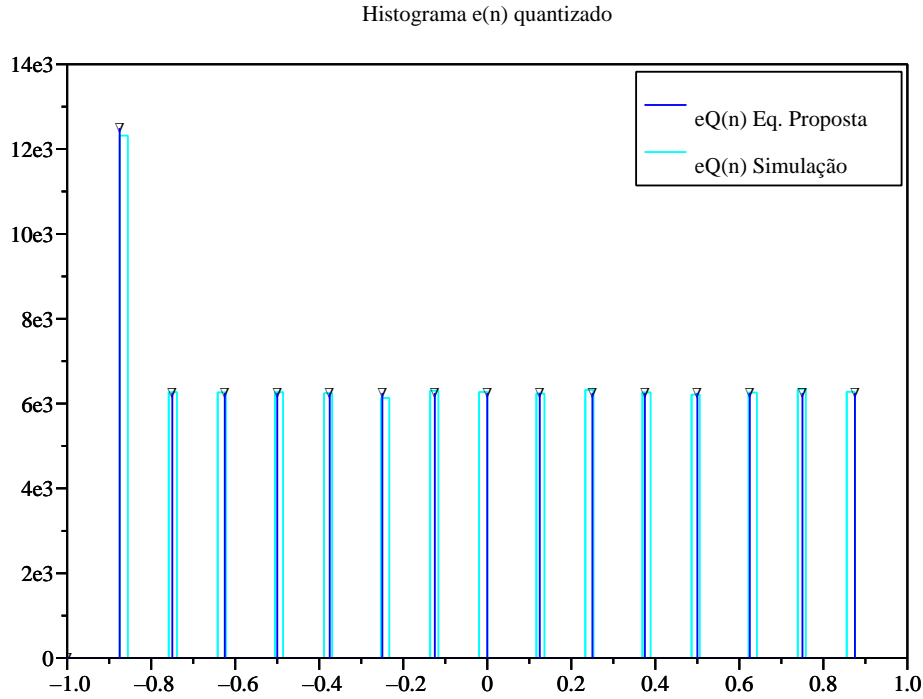


Figura 3.9: Comparação entre o Modelo e a Simulação para $f_{e_Q}(e_Q)$ com $b = 4bits$, $\sigma_z = 10^{-4}$, $\hat{w} = 0$ e $w(0) = -1$

3.5.3 Exemplos para $f_w(w)$

A seguir apresentam-se histogramas comparativos para $f_{w(w)}$ obtidos pelo modelo teórico proposto e obtidos pela simulação. A atualização do coeficiente do filtro quantizado é definido pela expressão $w(n+1) = Q\{w(n) + \mu e_Q(n)x(n)\}$ em que os coeficientes do filtro, o erro e o sinal de entrada encontram-se quantizados em b bits. As Figuras 3.10, 3.11, 3.12 e 3.13 mostram as probabilidades de $w(n)$ obtidas pelas simulações Monte Carlo e os resultados previstos pelas eqs. (3.21) e (3.22). As simulações executadas consideram a média de 100000 realizações. Os exemplos consideram diferentes valores de: passo de atualização μ ; coeficiente ótimo \hat{w} ; e condição inicial para o coeficiente do filtro $w(n)$. Nas quatro figuras o modelo e as simulações apresentam uma muito boa combinação, caracterizando uma excelente acurácia do modelo proposto para as probabilidades de $w(n)$ quantizado.

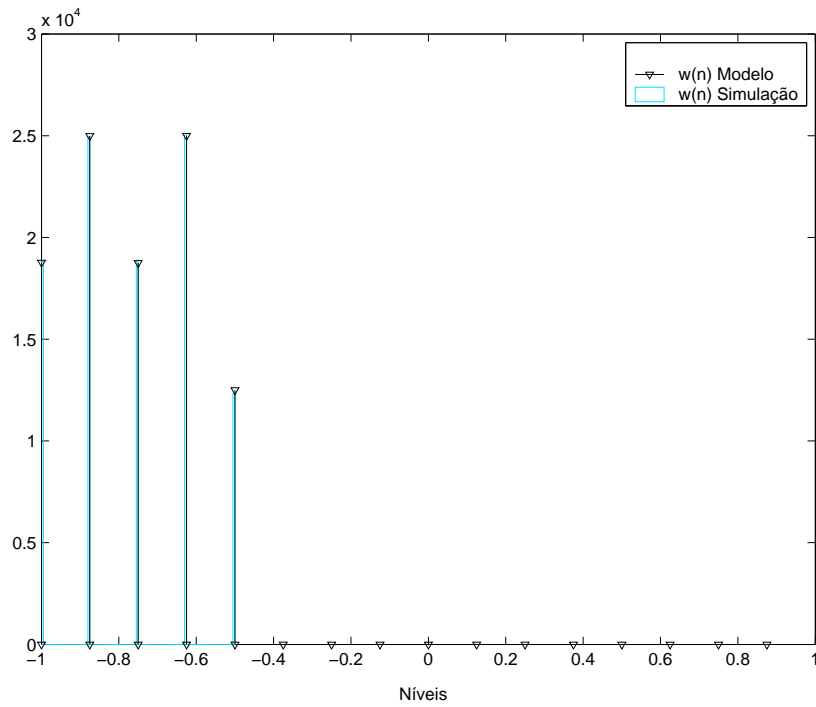


Figura 3.10: Simulação versus equações (3.21) e (3.22), para $b = 4$ bits, $\mu = 0,5$, $\hat{w} = 0,875$, $w(0) = -1$ e $\sigma = 10^{-3}$

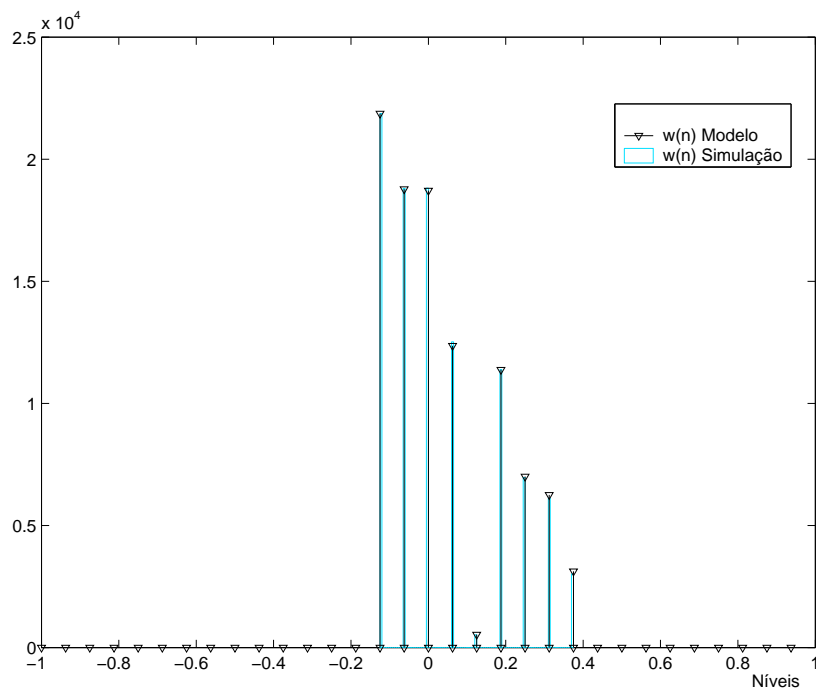


Figura 3.11: Simulação versus equações (3.21) e (3.22), $b = 5$ bits, $\mu = 0,7$, $\hat{w} = 0,5625$, $w(0) = -0,125$ e $\sigma = 10^{-3}$

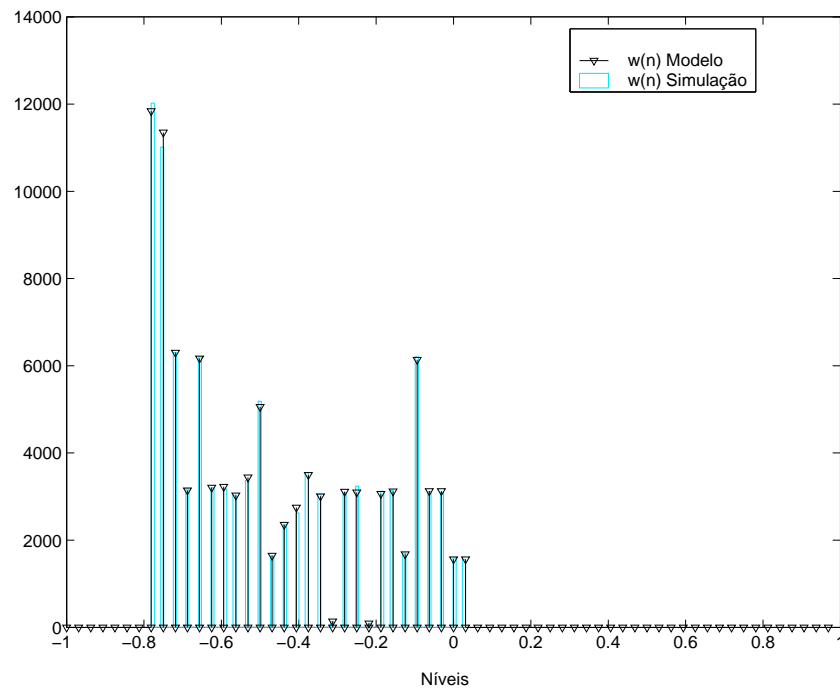


Figura 3.12: Simulação versus equações (3.21) e (3.22), para $b = 6$ bits, $\mu = 0,8$, $\hat{w} = 0,40625$, $w(0) = -0,78125$ e $\sigma = 10^{-3}$

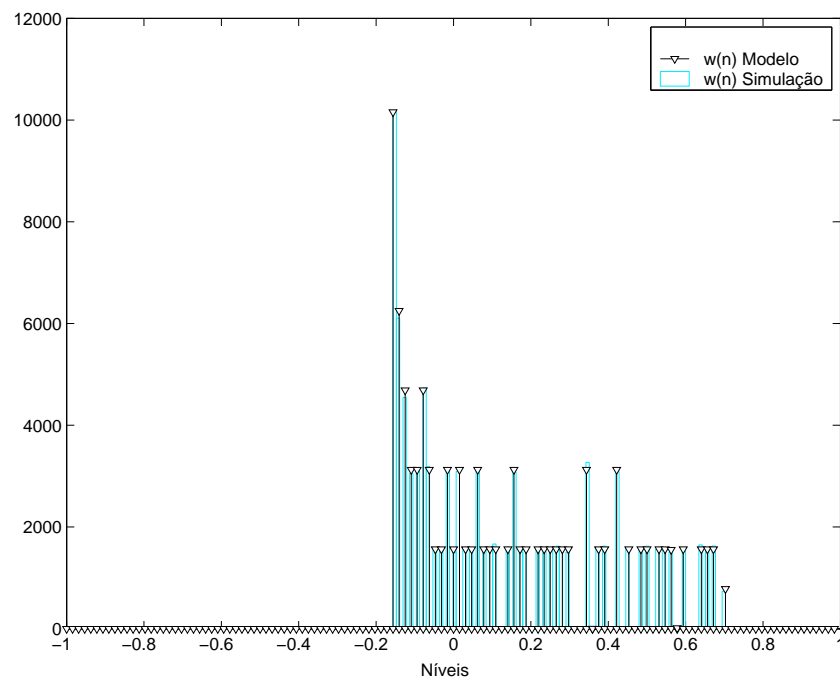


Figura 3.13: Simulação versus equações (3.21) e (3.22), para $b = 7$ bits, $\mu = 1$, $\hat{w} = 0,703125$, $w(0) = -0,15625$ e $\sigma = 10^{-5}$

3.5.4 Exemplos para Matriz de probabilidade para $w(n)$, P_w

O modelo proposto dado pelas eqs. (3.21) e (3.22) tem sido testado em várias situações diferentes. Nesta seção apresentam-se algumas amostras destes resultados para ilustrar a acurácia do modelo. A seguir exemplos da matriz de probabilidade P_w gerada por simulação e pelo modelo proposto.

Exemplo 1: Consideram-se $b = 2$ bits, $\hat{w} = 0,5$, $\mu = 0,5$ e $\sigma = 10^{-3}$. Os níveis de quantização são dados por $\{w(n) \in [-1 \ -0,5 \ 0 \ 0,5]\}$. A matriz P_w determinada por simulação representa a média de 100.000 realizações.

a) *Simulação*

$$\begin{array}{c} \textbf{Estados} \\ \left[\begin{array}{c} -1,00 \\ -0,50 \\ 0,00 \\ 0,50 \end{array} \right] \end{array} \quad P_w = \begin{pmatrix} 0,625 & 0,375 & 0 & 0 \\ 0 & 0,627 & 0,373 & 0 \\ 0 & 0 & 0,751 & 0,249 \\ 0 & 0 & 0 & 1,000 \end{pmatrix}$$

b) *Modelo Proposto*

$$\begin{array}{c} \textbf{Estados} \\ \left[\begin{array}{c} -1,00 \\ -0,50 \\ 0,00 \\ 0,50 \end{array} \right] \end{array} \quad P_w = \begin{pmatrix} 0,625 & 0,375 & 0 & 0 \\ 0 & 0,625 & 0,375 & 0 \\ 0 & 0 & 0,750 & 0,250 \\ 0 & 0 & 0 & 1,000 \end{pmatrix}$$

Comentários: Pode-se observar a grande semelhança que existe entre as probabilidades obtidas por simulação e as obtidas pelo modelo teórico proposto definido pelas eqs. (3.21) e (3.22).

Exemplo 2: Neste exemplo, $\hat{w} = -0,25$, $\mu = 1$, $\sigma = 10^{-3}$ e $b = 3$ bits. Os níveis de quantização para $w(n)$ são: $w = [-1 \ -0,75 \ -0,5 \ -0,25 \ 0 \ 0,25 \ 0,5 \ 0,75]$. Novamente, consideram-se 100.000 realizações para P_w determinada por simulação.

a) *Simulação*

$$\begin{array}{c} \mathbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$P_w = \begin{pmatrix}
 \begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} &
 \begin{array}{cccccccc}
 0,37 & 0,25 & 0,38 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0,38 & 0,50 & 0,12 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0,50 & 0,50 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,87 & 0,13 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,50 & 0,50 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,37 & 0,25 & 0,38 & 0 & 0 \\
 0 & 0 & 0 & 0,12 & 0,25 & 0,25 & 0,38 & 0 \\
 0 & 0 & 0 & 0 & 0,13 & 0,38 & 0,12 & 0,37
 \end{array}
 \end{pmatrix}$$

b) *Modelo Proposto*

$$\begin{array}{c} \mathbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$P_w = \begin{pmatrix}
 \begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} &
 \begin{array}{cccccccc}
 0,37 & 0,25 & 0,38 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0,38 & 0,50 & 0,12 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0,50 & 0,50 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,87 & 0,13 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,50 & 0,50 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,38 & 0,25 & 0,37 & 0 & 0 \\
 0 & 0 & 0 & 0,13 & 0,25 & 0,25 & 0,37 & 0 \\
 0 & 0 & 0 & 0 & 0,13 & 0,38 & 0,12 & 0,37
 \end{array}
 \end{pmatrix}$$

Comentários: Pode-se observar que existe uma excelente combinação entre a simulação e o modelo teórico derivado.

a) *Modelo Proposto*

$$Estados = \begin{bmatrix} -1 & -0,875 & -0,75 & -0,625 & -0,5 & -0,375 & -0,25 & -0,125 & 0 & 0,125 & 0,25 & 0,375 & 0,5 & 0,625 & 0,75 & 0,875 \end{bmatrix}$$

$$P_w = \begin{pmatrix} \begin{matrix} -1 \\ -0,875 \\ -0,75 \\ -0,625 \\ -0,5 \\ -0,375 \\ -0,25 \\ -0,125 \\ 0 \\ 0,125 \\ 0,25 \\ 0,375 \\ 0,5 \\ 0,625 \\ 0,75 \\ 0,875 \end{matrix} & \begin{matrix} 0,188 & 0,250 & 0,188 & 0,250 & 0,125 & 0 & 0 & 0 & 0,125 & 0,188 & 0,250 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0,188 & 0,250 & 0,188 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0,125 & 0,188 & 0,250 & 0,125 & 0 & 0 \\ 0 & 0 & 0,188 & 0,250 & 0,125 & 0,188 & 0,250 & 0,125 & 0 & 0 & 0 & 0 & 0,125 & 0,188 & 0,250 & 0,125 \\ 0 & 0 & 0 & 0,312 & 0,125 & 0,246 & 0,188 & 0,191 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,312 & 0,188 & 0,184 & 0,191 & 0,125 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,312 & 0,188 & 0,188 & 0,188 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,312 & 0,250 & 0,250 & 0,125 & 0,188 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,320 & 0,320 & 0,243 & 0,250 & 0,125 & 0,063 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,320 & 0,360 & 0,132 & 0,125 & 0,062 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,320 & 0,250 & 0,063 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,375 & 0,125 & 0,062 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,375 & 0,500 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,618 & 0,320 & 0,062 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,805 & 0,195 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,937 & 0,063 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix} \end{pmatrix}$$

Comentários: Da mesma forma que os exemplos anteriores, este exemplo valida os resultados do modelo proposto confirmando a previsão teórica e a excelente acurácia do modelo proposto.

Deve destacar-se que o tempo de processamento computacional para a determinação da matriz P_w usando o modelo proposto é significativamente menor do que o tempo necessário para a simulação.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,33 & 0,67 \end{pmatrix}$$

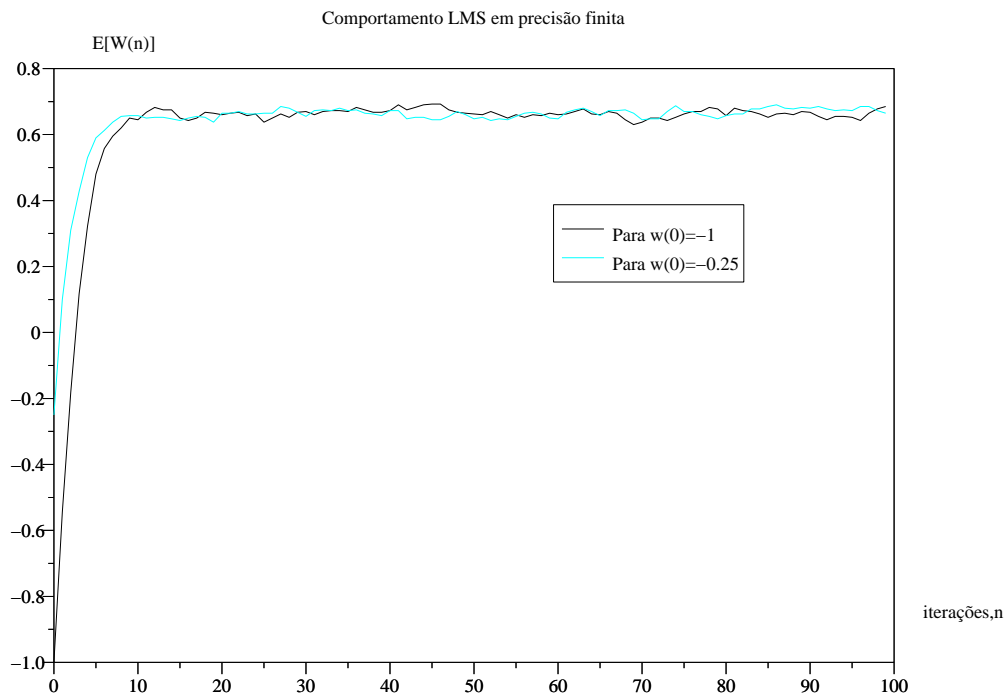


Figura 3.14: Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 1$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = -0,25$

Comentários: Observe que na matriz de probabilidade P_w^{100} quando o estado presente de $w(0) = -1$, existe 33% de probabilidade de que $w(n)$ convirja para o valor 0,5 e 67% de probabilidade de que convirja para o valor 0,75. O mesmo acontece quando o estado inicial é $w(0) = -0,25$. Assim, na média o próximo estado em regime permanente é $w(1) = 0,668$. Compare este resultado com a magnitude alcançada em $E[w(n)]$, Figura 3.14. Pode observar-se que o $E[W(n)]$ em regime permanente previsto pela matriz P_w^{100} ajusta-se perfeitamente aos resultados das simulações.

Observe também como o comportamento da matriz P_w^{100} obedece a uma matriz quase-ergódica conforme a classificação das cadeias de Markov de ordem 1 e as conclusões do Capítulo 2.

Exemplo 2: Neste exemplo considera-se: um quantizador de $b = 3$ bits; condição inicial para o coeficiente do filtro de $\hat{w} = 0,75$; sinal de ruído com $\sigma_z = 10^{-3}$; e um passo de adaptação de $\mu = 0,58$. P_w é a matriz para uma transição, P_w^{100} é matriz para $n = 100$ transições composta pelas probabilidades estacionárias. Na Figura 3.15 mostra-se o comportamento de $E[W(n)]$ para o algoritmo LMS em precisão finita. Apresentam-se duas curvas, uma considerando $w(0) = -1$ e a outra com $w(0) = 0,25$. Considera-se para as a simulações de Monte Carlo a média de 100 realizações.

$$\begin{array}{c} \mathbf{Estados} \end{array} \quad \left[\begin{array}{cccccccc} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{array} \right]$$

$$P_w = \left(\begin{array}{cccccccc} 0,38 & 0,37 & 0,25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,38 & 0,37 & 0,25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,38 & 0,37 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,37 & 0,50 & 0,13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,50 & 0,37 & 0,13 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,87 & 0,13 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,87 & 0,13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

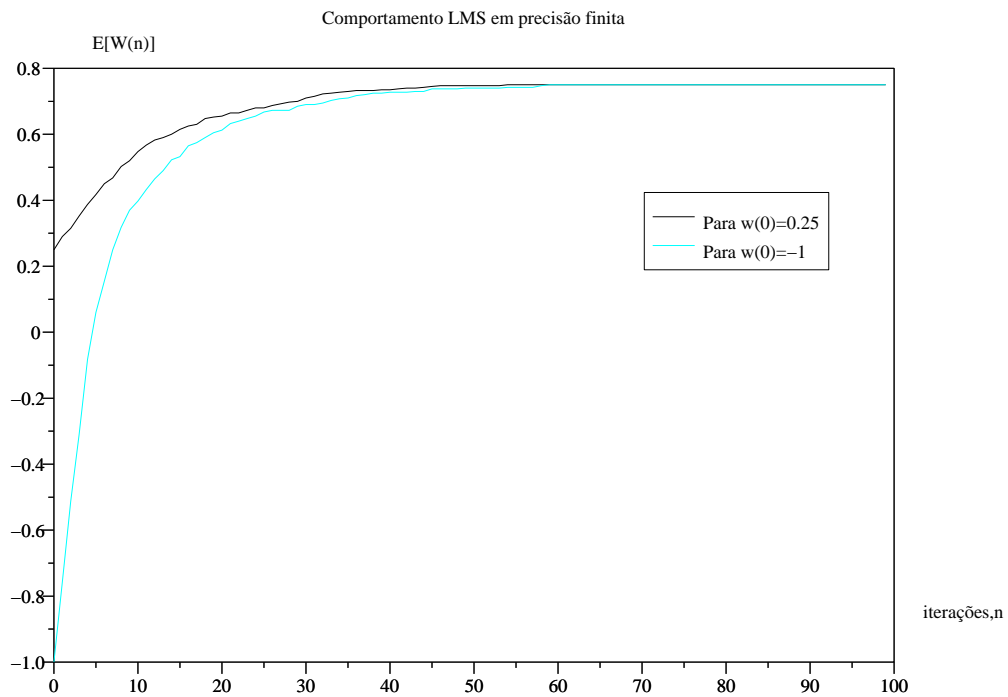


Figura 3.15: Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 0,58$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = 0,25$

Comentários: Observe que na matriz de probabilidade P_w^{100} quando o estado inicial é $w(0) = -1$, existe 100% de probabilidade de que $w(n)$ convirja para o valor 0,75, o mesmo acontece com $w(0) = 0,25$. Compare este resultado com a magnitude alcançada para $E[w(n)]$ na Figura 3.15. Pode-se observar que o $E[W(n)]$ em regime permanente previsto por P_w^{100} coincide perfeitamente com os resultados obtidos por simulação.

Para este exemplo a matriz P_w^{100} também é uma matriz quase-ergódica, em que independente do estado presente o próximo estado de $w(n)$ sempre será o coeficiente ótimo $\hat{w} = 0,75$.

Exemplo 3: Neste exemplo considera-se: um quantizador de $b = 3$ bits; condição inicial para o coeficiente do filtro de $\hat{w} = 0,75$; sinal de ruído com $\sigma_z = 10^{-3}$; e um passo de adaptação de $\mu = 0,2$. P_w é a matriz para uma transição, P_w^{100} é matriz para $n = 110$ transições composta pelas probabilidades estacionárias. Na Figura 3.16 mostra-se o comportamento de $E[W(n)]$ para o algoritmo LMS em precisão finita. Apresentam-se duas curvas, uma considerando $w(0) = -1$ e a outra com $w(0) = 0,75$. Considera-se para as simulações de Monte Carlo a média de 110 realizações.

$$\begin{array}{c} \mathbf{Estados} \end{array} \quad \left[\begin{array}{cccccccc} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{array} \right]$$

$$P_w = \begin{pmatrix} 0,75 & 0,25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,75 & 0,25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,75 & 0,25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,87 & 0,13 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,87 & 0,13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1,00 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \end{pmatrix}$$

$$\begin{array}{c} \text{Estados} \end{array} \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$P_w^{110} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

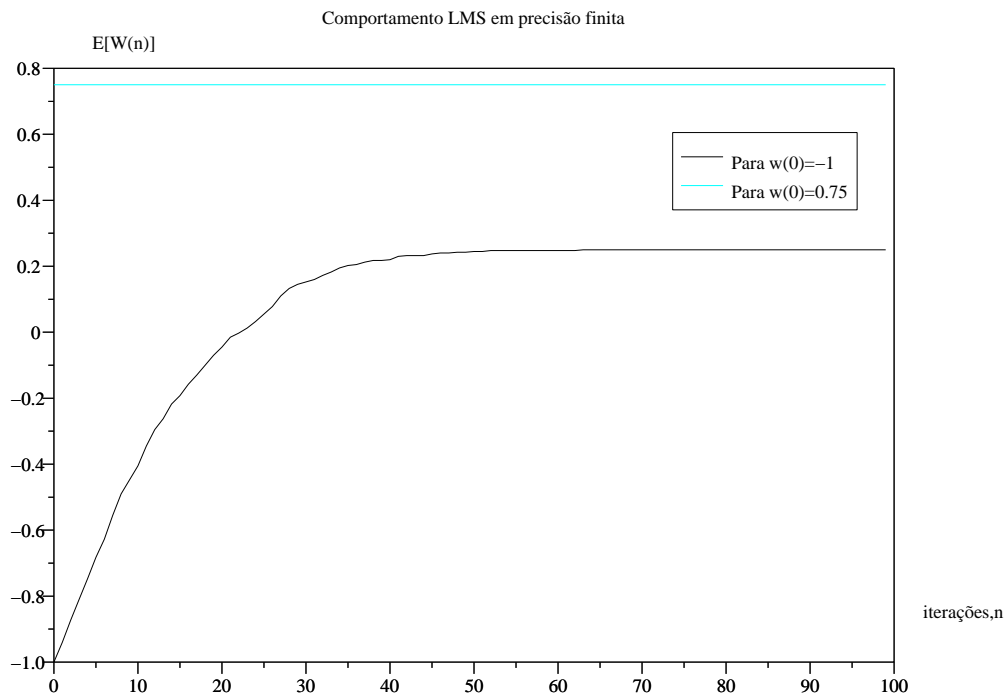


Figura 3.16: Algoritmo LMS em precisão finita com $b = 3$ bits, $\mu = 0,2$, $\sigma = 10^{-3}$ e $\hat{w} = 0,75$, para $w(0) = -1$ e $w(0) = 0,75$

Comentários: Observe que na matriz de probabilidade P_w^{110} , quando o estado inicial é $w(0) = -1$, existe 100% de probabilidade de que $w(n)$ convirja para o valor 0,25 e quando o estado inicial é $w(0) = 0,75$ existe 100% de probabilidade que $w(n)$ convirja para o mesmo valor, 0,75. Compare estes resultados com as magnitudes alcançadas para $E[W(n)]$. Pode-se observar comparando os resultados obtidos pelo modelo (P_w^{110}) e por simulação, a excelente acurácia do modelo proposto.

Para este exemplo a matriz P_w^{110} é uma matriz não ergódica, observe que o estado futuro de $w(n)$ dependente do estado presente, isto devido a que a magnitude do passo μ não permite a mudança de estado para $w(n) = 0,25, 0,5$ e $0,75$, fazendo que na matriz P_w existam três estados absorventes, assim, no futuro os demais estados de $w(n)$ terminarão sendo absorvidos por alguns destes três estados.

3.6 CONCLUSÕES

- O modelo proposto para a fdp de $w(n)$ confirma o desempenho previsto pelo modelo teórico, caracterizando a acurácia do modelo proposto.
- O tempo utilizado no processamento para a determinação da matriz P_w é significativamente menor do que o tempo necessário para a simulação desta matriz, demonstrando a eficácia do modelo derivado.
- Observa-se da Seção 3.5.5 que a convergência do algoritmo LMS em precisão finita ao coeficiente ótimo \hat{w} depende da escolha do tamanho do passo μ . Para o Exemplo 2, a presença de um estado absorvente em P_w quando o estado presente é igual a \hat{w} assegura que o próximo estado de $w(n)$ é o coeficiente ótimo com 100% de probabilidade. Assim, em regime permanente o algoritmo permanece neste estado independente do estado inicial.

Capítulo 4

LIMITES DO PASSO DE ADAPTAÇÃO

4.1 INTRODUÇÃO

Conhecidas as probabilidades de transição entre os estados dos coeficientes, é possível determinar condições a serem satisfeitas para que P_w tenha propriedades desejáveis para o comportamento adequado do algoritmo LMS em precisão finita. Por exemplo, o algoritmo deve ser projetado de forma a permitir a convergência do coeficiente para o seu valor ótimo \hat{w} na ausência de ruído de medição, independentemente do valor do estado inicial de $w(n)$. Esse requisito implica que a matriz P_w deverá ter uma única classe final, com um único estado absorvente em $w(n) = \hat{w}$. Assim, P_w deve corresponder a uma cadeia de Markov quase-ergódica [15].

Inicialmente, existe a possibilidade de paralização do algoritmo em um ambiente de precisão finita. Basta para isso que o ajuste gerado pelo algoritmo seja menor que a metade de um passo de quantização: neste caso $w(n+1) = w(n)$ em (3.3) e o algoritmo ficará paralizado. Outro aspecto importante é a garantia de que $w(n) = \hat{w}$ seja um estado absorvente da cadeia de Markov, para que \hat{w} seja a solução em regime permanente. Esses dois aspectos são estudados neste capítulo.

Em situações reais de aplicação do algoritmo LMS o ruído de medição é considerável, alterando os resultados ideais. As propriedades estatísticas do ruído de medição podem al-

terar a probabilidade de que $\lim_{n \rightarrow \infty} w(n) = \hat{w}$, mesmo que as condições de operação ideal sem ruído sejam satisfeitas.

A última seção deste capítulo apresenta uma análise da influência da estatística do ruído de medição na probabilidade de convergência do algoritmo para o coeficiente de Wiener.

4.2 O DIAGRAMA DE TRANSIÇÃO DOS ESTADOS DE

$$w(n)$$

Nesta seção apresentam-se três exemplos da matriz de probabilidades de transição de estados P_w para diferentes valores do passo de adaptação. Analisam-se então as propriedades do diagrama de transição em cada caso, determinando o tipo de cadeia de Markov que representa o comportamento do algoritmo LMS em precisão finita.

Exemplo 1: Considere $b = 3$ bits, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 1$. Neste caso a matriz P_w é dada por:

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w = \begin{pmatrix}
 \begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} & \begin{bmatrix}
 0,125 & 0,250 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\
 0 & 0,125 & 0,250 & 0,375 & 0,250 & 0 & 0 & 0 \\
 0 & 0 & 0,125 & 0,375 & 0,250 & 0,250 & 0 & 0 \\
 0 & 0 & 0 & 0,375 & 0,125 & 0,375 & 0,125 & 0 \\
 0 & 0 & 0 & 0 & 0,375 & 0,250 & 0,250 & 0,125 \\
 0 & 0 & 0 & 0 & 0 & 0,500 & 0,375 & 0,125 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0,750 & 0,250 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0,125 & 0,875
 \end{bmatrix}
 \end{pmatrix}$$

O diagrama de transição de estados correspondente a essa matriz de probabilidades P_w é apresentado na Figura 4.1. Na Figura 4.1 omitiram-se os valores das probabilidades que relacionam os estados, para observar melhor algumas propriedades deste diagrama relativas

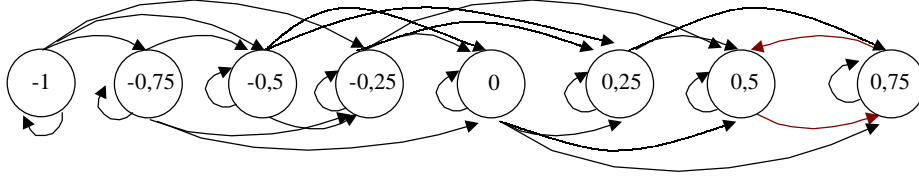


Figura 4.1: Diagrama de transição de estados de $w(n)$ com $b = 3\text{bits}$, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 1$

às possíveis transições de estados. Observa-se a existência de uma classe final composta pelos estados, 0,5 e 0,75 e de seis classes de passos: -1 , $-0,75$, $-0,5$, $-0,25$, 0 e $0,25$. O algoritmo LMS em precisão finita leva a mudança de estados até cair em algum dos estados da classe final, não voltando a sair mais dessa classe. Em estado estacionário, para cada um de seus estados de $w(n)$ permanecerá dentro dessa classe com certa probabilidade de ocorrência. Como estudou-se no Capítulo 3, basta então considerar os estados da classe final para calcular as probabilidades de estado estacionário (regime permanente). Assim, pode-se estudar apenas a matriz

$$P_{w1} = \begin{pmatrix} 0,750 & 0,250 \\ 0,125 & 0,875 \end{pmatrix}$$

Para esta matriz P_{w1} , correspondente às colunas e linhas P_w relativas aos estados 0,5 e 0,75, as probabilidades de estado estacionário obtidas do $\lim_{m \rightarrow \infty} P_{w1}^m$ são:

$$\pi_1 = 0,333$$

$$\pi_2 = 0,667$$

Isso significa que, em regime permanente qualquer estado de $w(n)$ terá uma probabilidade de 33% de cair no estado 0,5 e de 67% de cair no estado 0,75. Observe a matriz P_w para 100 passos no Capítulo 3, Seção 3.5.5, *Exemplo 1*.

Exemplo 2: Considere $b = 3$ bits, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,58$. A matriz P_w será:

$$\begin{array}{c} \textbf{Estados} \end{array} \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w = \begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,500 & 0,375 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 \end{pmatrix}$$

O diagrama de transição de estados para este exemplo é mostrado na Figura 4.2,

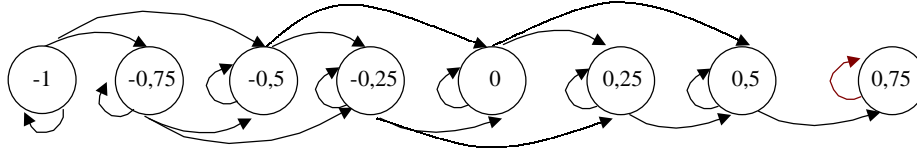


Figura 4.2: Diagrama de transição de estados de $w(n)$ com $b = 3\text{bits}$, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,58$

Como no exemplo anterior o diagrama de transição não contém os valores das probabilidades de transição para facilitar a observação das classes. Observa-se que existe uma classe final composta por um estado absorvente ($P = 1$) no estado 0,75, sendo que os demais estados formam classes de passos. O algoritmo LMS em precisão finita levará $w(n)$ ao estado 0,75, tendo 100% de probabilidade de permanecer nesse estado. Em regime permanente, todos os estados de passo terminarão sendo absorvidos por este estado da classe final. Observe a matriz P_w após 100 passos em *Exemplo 2*, Seção 3.5.5, Capítulo 3. A cadeia de Markov é quase-ergódica, o algoritmo converge para $\hat{w} = 0,75$ com 100% de probabilidade.

Exemplo 3: Considere $b = 3$ bits, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,2$. P_w será

$$\text{Estados} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w = \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \\ 0,750 & 0,250 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,750 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,750 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,875 & 0,125 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1,000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 \end{bmatrix}$$

O diagrama de transição de estados para P_w é dado na Figura 4.3.

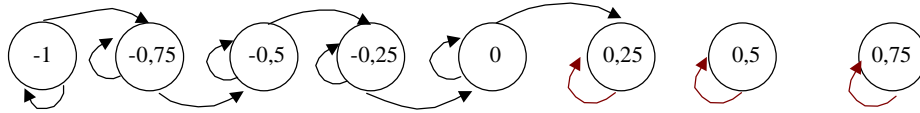


Figura 4.3: Diagrama de transição de estados de $w(n)$ com $b = 3$ bits, $\hat{w} = 0,75$, $\sigma_z = 10^{-3}$ e $\mu = 0,2$

Observa-se para este exemplo a existência de três classes finais nos estados $\{0,25, 0,5, 0,75\}$ (elementos iguais a 1,000 em P_w). Cada um deles é estado absorvente. Os estados de passos $-1, -0,75, -0,5, -0,25, 0$ terminarão sendo absorvidos com 100% de probabilidade pelo estado 0,25. Observe no Capítulo 3, seção 3.5.5, *Exemplo 3* a matriz P_w após 110 passos. Observe que trata-se de uma matriz não ergódica, em que a probabilidade do estado estacionário dependerá do estado inicial.

Das análises anteriores tem-se, que o *Exemplo 2* representa o comportamento ideal do algoritmo LMS em precisão finita. Em regime permanente e independentemente do estado inicial para $w(n)$, o algoritmo sempre terminará convergindo para o coeficiente ótimo \hat{w} .

Observe que para este *Exemplo* a matriz P_w é uma matriz quase-ergódica, com uma classe final que contém um único estado (estado absorvente) igual a \hat{w} . A partir destas observações determina-se a seguir as condições sobre o passo μ para que o coeficiente convirja para estado ótimo independentemente do estado inicial de $w(n)$.

4.3 PASSO DE ADAPTAÇÃO- GARANTIA DE ESTADOS ABSORVENTES

Deseja-se determinar μ tal que a matriz P_w tenha um estado absorvente. Primeiro determina-se as condições sobre o passo para que existam N_w estados absorventes.

Retomando a equação de atualização do algoritmo LMS para entradas e coeficientes discretos:

$$w(n+1) = w(n) + y_Q(n) \quad (4.1)$$

Tem-se que:

Se $P\{y_Q(n) = 0\}$ é igual a 1 para todo $w(n)$ e $x(n)$, então ocorrerá sempre que $w(n+1) = w(n)$, e a matriz de probabilidades de transição P_w terá mais de um estado absorvente. Essa condição é equivalente a:

$$\begin{aligned} & P\{y_Q(n) = 0\} = 1 \\ \text{ou} \quad & P\left\{\frac{-\Delta}{2} < y(n) < \frac{\Delta}{2}\right\} = 1 \\ \text{ou} \quad & P\left\{\frac{-\Delta}{2} < \mu e_Q(n)x(n) < \frac{\Delta}{2}\right\} = 1, \quad \text{para todo } x(n) \end{aligned}$$

Para determinar um valor de μ , tal que $y_Q(n) = 0$ para qualquer estado de entrada $x(n)$ e para qualquer coeficiente $w(n)$, deve considerar-se $|e_Q(n)|$ e $|x(n)|$ com valores máximos,

ou seja, $|e_Q(n)| = |x(n)| = 1$. Neste caso, considerando que $\mu > 0$, tem-se:

$$\mu < \frac{\Delta}{2} \quad (4.2)$$

Portanto com $\mu < \frac{\Delta}{2}$, a matriz P_w será igual à matriz identidade, $P_w = I$. Observe a matriz P_w para $b = 3$ bits, $\hat{w} = -0,25$, $\sigma_z = 10^{-3}$ e $\mu = 0,124$

$$Estados = \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$P_w = \begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

4.4 LIMITE INFERIOR DO PASSO DE ADAPTAÇÃO

Para evitar paralização do algoritmo, o termo de ajuste do coeficiente em (3.3) deve satisfazer

$$|y_Q(n)| \geq \Delta$$

ou seja,

$$|Q\{ \mu e_Q(n)x(n) \}| \geq \Delta \quad (4.3)$$

ou equivalentemente,

$$|\mu e_Q(n)x(n)| \geq \frac{\Delta}{2} \quad (4.4)$$

Como $|e_Q(n)| \geq \Delta$ (para ruído baixo) sempre que $w(n) \neq \hat{w}$, (4.4) resulta na condição $|\mu x(n)| \geq \frac{1}{2}$. O mínimo valor de μ que satisfaz essa condição ocorre para $|x(n)| = 1$. Assim, a condição sobre o passo μ para que o algoritmo não fique paralizado para $w(n) \neq \hat{w}$ é

$$\mu \geq \frac{1}{2} \quad (4.5)$$

4.5 LIMITE SUPERIOR DO PASSO DE ADAPTAÇÃO

Para que $w(n) = \hat{w}$ seja um estado absorvente da cadeia de Markov, $y_Q(n)$ deve ser igual a zero para toda entrada $x(n)$ quando $w(n) = \hat{w}$. Desta forma teremos que $w(n+1) = w(n)$. Essa condição requer

$$Pr\{y_Q(n) = 0\} = 1$$

ou seja,

$$Pr\left\{y(n) < \frac{\Delta}{2}\right\} = 1$$

ou

$$\left\{(\mu e_Q k_x \Delta) < \frac{\Delta}{2}\right\}$$

ou ainda

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{w(n)x(n)\}| \quad (4.6)$$

Analisando essas expressões, tem-se que:

1. Com $|x(n)| = 1$ a equação (4.6) fica,

$$|e_Q(n)| = |Q\{\hat{w} + z(n)\} - Q\{w(n)\}|$$

para $|z(n)| \ll \frac{\Delta}{2}$ tem-se que,

$$|e_Q(n)| \cong |Q\{\hat{w}\} - Q\{w(n)\}| = 0$$

Assim, $y_Q(n) = 0$ para $|x(n)| = 1$

2. Com $|x(n)| \neq 1$ sempre poderá acontecer que $e_Q(n) \neq 0$, pois, por exemplo, se $\hat{w}x(n) = k\frac{\Delta}{2}$,¹

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{w(n)x(n)\}|$$

e,

$$|e_Q(n)| = \left| Q\left\{k\frac{\Delta}{2} + z(n)\right\} - Q\left\{k\frac{\Delta}{2}\right\} \right|$$

para $|z(n)| \ll \frac{\Delta}{2}$ mas não nulo, o pior caso para o estado de erro é

$$|e_Q(n)| = \Delta \quad (4.7)$$

Substituindo (4.7) na expressão de $y(n)$ tem-se:

$$|y_Q(n)| = |Q\{\mu\Delta x(n)\}|$$

¹ Assumindo que o valor \hat{w} coincida com um estado w_n . Se \hat{w} não satisfizer essa condição, estaremos determinando o limite para μ que garante que w_n convirja para o estado mais próximo de \hat{w} .

Para que $y_Q(n) = 0$, devemos ter $|y(n)| < \frac{\Delta}{2}$. Assim,

$$|\mu x(n)| < \frac{1}{2}$$

O pior caso para $|x(n)| \neq 1$ é $|x(n)| = 1 - \Delta$ (estado menor seguinte a $|x(n)| = 1$).

Logo a condição fica

$$\mu < \frac{1}{2(1 - \Delta)}$$

Como $\Delta = 2^{1-b}$, tem-se,

$$\mu < \frac{2^b}{2^{b+1} - 4}$$

Assim no intervalo,

$$\boxed{\frac{1}{2} \leq \mu < \frac{2^b}{2^{(b+1)} - 4}} \quad (4.8)$$

a matriz P_w será quase-ergódica com um estado absorvente em $w(n) = \hat{w}$ para todo estado de $x(n)$.

Note que para valores práticos de b em que $b \geq 2$, o passo de adaptação pode ser representado com a mesma precisão numérica que os sinais e os coeficientes, já que $\mu = 0,5$ sempre será um nível de quantização.

4.6 Exemplos

Desenvolvendo (4.8) para alguns valores de b pode-se determinar os limites do passo μ para diferentes valores de b . Esses resultados estão apresentados na Tabela 4.1.

Tabela 4.1: Limites do passo μ para diferentes valores de b

b (bits)	μ
2	$0,5 \leq \mu < 1,000$
3	$0,5 \leq \mu < 0,670$
4	$0,5 \leq \mu < 0,570$
5	$0,5 \leq \mu < 0,533$
6	$0,5 \leq \mu < 0,516$
7	$0,5 \leq \mu < 0,507$

A seguir ilustram-se alguns exemplos da matriz P_w , e as probabilidades de estado estacionário P_w^{1000} , para diferentes valores do passo de adaptação. Apresentam-se dois casos:

4.6.1 Caso 1:

Considere $\hat{w} = -0,5$, $b = 3$ bits, $\sigma_z = 10^{-3}$. Neste caso serão considerados quatro valores de μ : 0,1, 0,3, 0,5 e 0,7, observe que somente $\mu = 0,5$ está dentro da faixa estipulada na Tabela 4.1.

- **Caso 1a:** $\hat{w} = -0,5$ $b = 3 \text{ bits}$ $\sigma_z = 10^{-3}$, $\mu = 0,1$

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 \end{pmatrix}$$

Sendo o passo de adaptação menor que $\frac{\Delta}{2}$ (0,125) a matriz P_w é igual à matriz identidade.

Observe a seguir as probabilidades em regime estacionário.

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{1000} = \begin{pmatrix} 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 \end{pmatrix}$$

- **Caso 1b:** $\hat{w} = -0,5$ $b = 3 \text{ bits}$ $\sigma_z = 10^{-3}$, $\mu = 0,3$

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 1,000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1,000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1,000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,125 & 0,875 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,250 & 0,750 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,375 & 0,625 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,375 & 0,625 \end{pmatrix}$$

Observe como com um passo superior a $\frac{\Delta}{2}$ a quantidade de estados absorventes em P_w diminui, mas a cadeia segue sendo não ergódica, veja a seguir o que acontece com as probabilidades em regime estacionário para o coeficiente adaptativo.

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{1000} = \begin{pmatrix} \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & \boxed{1,00} & 0,000 & 0,000 & 0,000 & 0,000 \end{pmatrix}$$

As probabilidades estacionárias em P_w^{1000} amostram a presença de três estados absorventes nos estados -1 , $-0,75$, $-0,5$ e $-0,25$.

- **Caso 2b:** $\hat{w} = 0,625$ $b = 4 \text{ bits}$ $\sigma_z = 10^{-4}$, $\mu = 0,4$

$$\begin{array}{c}
 \text{Estados} \\
 \left[\begin{array}{c} -1 \\ -0,875 \\ -0,75 \\ -0,625 \\ -0,5 \\ -0,375 \\ -0,25 \\ -0,125 \\ 0 \\ 0,125 \\ 0,25 \\ 0,375 \\ 0,5 \\ 0,625 \\ 0,75 \\ 0,875 \end{array} \right]
 \end{array}
 P_w =
 \begin{array}{c}
 \left(\begin{array}{cccccccccccccccccccc}
 0,312 & 0,250 & 0,313 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0,312 & 0,250 & 0,313 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0,312 & 0,250 & 0,313 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,312 & 0,250 & 0,313 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0,312 & 0,250 & 0,313 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,313 & 0,125 & 0,250 & 0,063 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,313 & 0,125 & 0,250 & 0,063 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,313 & 0,125 & 0,250 & 0,063 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,313 & 0,125 & 0,250 & 0,063 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,500 & 0,437 & 0,375 & 0,063 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,562 & 0,375 & 0,063 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,562 & 0,438 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,812 & 0,188 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,188 & 0,812
 \end{array} \right)
 \end{array}$$

- **Caso 2c:** $\hat{w} = 0,625$ $b = 4 \text{ bits}$ $\sigma_z = 10^{-4}$, $\mu = 0,56$

$$\begin{aligned}
 & \text{Estados} \quad \left[\begin{array}{cccccccccccccccccccc} -1 & -0,875 & -0,75 & -0,625 & -0,5 & -0,375 & -0,25 & -0,125 & 0 & 0,125 & 0,25 & 0,375 & 0,5 & 0,625 & 0,75 & 0,875 \end{array} \right] \\
 & P_w = \left(\begin{array}{cccccccccccccccccccc}
 -1 & 0,312 & 0,125 & 0,188 & 0,250 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -0,875 & 0 & 0,312 & 0,125 & 0,250 & 0,188 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -0,75 & 0 & 0 & 0,312 & 0,188 & 0,188 & 0,187 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -0,625 & 0 & 0 & 0 & 0,312 & 0,188 & 0,188 & 0,187 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -0,5 & 0 & 0 & 0 & 0 & 0,312 & 0,250 & 0,125 & 0,188 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -0,375 & 0 & 0 & 0 & 0 & 0 & 0,312 & 0,250 & 0,125 & 0,250 & 0,125 & 0 & 0 & 0 & 0 & 0 \\
 -0,25 & 0 & 0 & 0 & 0 & 0 & 0 & 0,312 & 0,375 & 0 & 0,250 & 0,063 & 0 & 0 & 0 & 0 \\
 -0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,250 & 0,063 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,250 & 0,063 & 0 & 0 & 0 \\
 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,500 & 0,437 & 0 & 0 & 0 & 0 \\
 0,25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,063 & 0 & 0 & 0 \\
 0,375 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,063 & 0 & 0 \\
 0,5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,187 & 0 & 0 \\
 0,625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,937 & 0,063 & 0 \\
 0,75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 \\
 0,875 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,937 & 0 \\
 & & & & & & & & & & & & & & 0,250 & 0,750
 \end{array} \right)
 \end{aligned}$$

- **Caso 2d:** $\hat{w} = 0,625$ $b = 4$ *bits* $\sigma_z = 10^{-4}$, $\mu = 1$

$$P_w = \begin{bmatrix} \text{Estados} & \begin{bmatrix} -1 & -0,875 & -0,75 & -0,625 & -0,5 & -0,375 & -0,25 & -0,125 & 0 & 0,125 & 0,25 & 0,375 & 0,5 & 0,625 & 0,75 & 0,875 \end{bmatrix} \\ \begin{bmatrix} -1 \\ -0,875 \\ -0,75 \\ -0,625 \\ -0,5 \\ -0,375 \\ -0,25 \\ -0,125 \\ 0 \\ 0,125 \\ 0,25 \\ 0,375 \\ 0,5 \\ 0,625 \\ 0,75 \\ 0,875 \end{bmatrix} & \begin{pmatrix} 0,187 & 0,125 & 0,125 & 0,063 & 0,125 & 0,125 & 0,125 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,187 & 0,125 & 0,250 & 0,063 & 0,187 & 0,063 & 0,125 & 0,125 & 0,125 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,187 & 0,125 & 0,125 & 0,125 & 0,125 & 0,063 & 0,125 & 0,125 & 0,125 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,187 & 0,125 & 0,125 & 0,125 & 0,063 & 0,125 & 0,125 & 0,063 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,187 & 0,125 & 0,187 & 0,063 & 0,125 & 0,125 & 0,125 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,187 & 0,250 & 0,063 & 0,187 & 0 & 0,125 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,187 & 0,250 & 0,125 & 0,125 & 0 & 0,250 & 0 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,313 & 0,125 & 0,125 & 0,125 & 0,250 & 0 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,313 & 0,187 & 0,187 & 0,125 & 0,125 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,313 & 0,375 & 0,125 & 0,125 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,437 & 0,250 & 0,250 & 0,063 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,500 & 0,3125 & 0,187 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,750 & 0,250 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,063 & 0,937 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,500 & 0,500 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,187 & 0,375 & 0,437 \end{pmatrix} \end{bmatrix}$$

Observe que para um passo de adaptação $\mu = 1$, acima do intervalo, não existem estados absorventes, mas a cadeia segue sendo quase-ergódica.

A seguir apresentam-se as probabilidades estacionárias, em P_w^{1000} . Observe que o estado 0,625 tem 80% de probabilidade de acontecer, enquanto que o estado 0,5 tem 20% de probabilidade. Assim, em regime permanente o algoritmo LMS independentemente do estado inicial do coeficiente, na média vai convergir para o valor 0,6.

$$\text{Estados} \quad \left[\begin{array}{cccccccccccccccc} -1 & -0,875 & -0,75 & -0,625 & -0,5 & -0,375 & -0,25 & -0,125 & 0 & 0,125 & 0,25 & 0,375 & 0,5 & 0,625 & 0,75 & 0,875 \end{array} \right] \quad P_w^{1000} = \left(\begin{array}{cccccccccccccccc} 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,200 & 0,800 & 0,000 & 0,000 \end{array} \right)$$

Todos os exemplos anteriores demonstram a validade da condição do passo de adaptação obtido.

4.7 PASSO DE ADAPTAÇÃO E POTÊNCIA DE RUÍDO

Na análise anterior considerou-se uma potência de ruído $z(n)$ o suficientemente pequena para não alterar os valores quantizados obtidos a partir dos sinais. Se a potência de ruído assumir valores significativos, o estado final do coeficiente passa a ter probabilidades não nulas de assumir um estado diferente de \hat{w} mesmo para valores bem pequenos do passo de adaptação.

Nesta seção estuda-se o comportamento em regime permanente de $w(n)$ no caso de ruído elevado

4.7.1 Limite inferior do passo de adaptação em função de σ

Para evitar a paralização do algoritmo, o termo de ajuste do coeficiente deve satisfazer

$$|Q\{\mu e_Q(n)x(n)\}| \geq \Delta \quad \text{quando } w(n) \neq \hat{w}$$

ou equivalentemente,

$$|\mu e_Q(n)x(n)| \geq \frac{\Delta}{2} \quad (4.9)$$

em que

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{w(n)x(n)\}|$$

Sabendo que o mínimo erro acontece quando $w(n) = \hat{w} \pm \Delta$ tem-se que,

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{(\hat{w} \pm \Delta)x(n)\}|$$

Para a análise a seguir considera-se,

$$|z(n)| < \frac{\beta\Delta}{2} \quad (4.10)$$

em que, $\frac{\beta\Delta}{2}$ é um ponto de decisão na quantização, portanto β é inteiro ímpar.

O mínimo valor de μ que satisfaz a condição (4.9) ocorre para $|x(n)| = 1$, então

$$|e_Q(n)| = |Q\{\hat{w} + z(n)\} - Q\{(\hat{w} \pm \Delta)\}| \quad (4.11)$$

Se $w(n) = \hat{w} + \Delta$ o pior caso para o erro acontece para valores negativos de $z(n)$. Observe a Figura 4.4 e o exemplo a seguir.

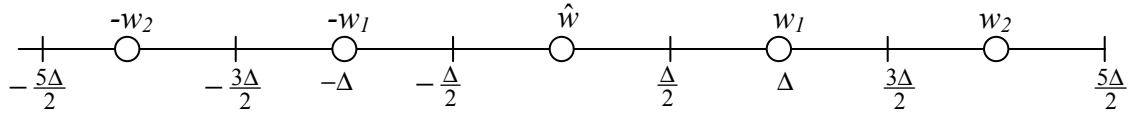


Figura 4.4: Ilustração do método de quantização

Exemplo: Para este exemplo considera-se $\beta = 3$, $w(n) = w_1$ e $|z(n)| = (\beta - \frac{1}{10^4})\frac{\Delta}{2}$. Sabe-se que $Q\{w\} = -w_1$ para $-\frac{3\Delta}{2} < w \leq -\frac{\Delta}{2}$ e $Q\{w\} = w_1$ para $\frac{\Delta}{2} \leq w < \frac{3\Delta}{2}$, então tem-se que:

Para $z(n)$ positivo, $|e_Q(n)| = |Q\{\hat{w} + z(n)\} - Q\{w_1\}| = |w_1 - w_1| = 0$. Enquanto que para $z(n)$ negativo, $|e_Q(n)| = |Q\{\hat{w} - z(n)\} - Q\{w_1\}| = |-w_1 - w_1| = 2w_1$.

Aplicando a análise anterior em (4.11), tem-se

$$\begin{aligned} |e_Q(n)| &= \left| \left(\hat{w} - \frac{(\beta-1)\Delta}{2} \right) - (\hat{w} + \Delta) \right| \\ |e_Q(n)| &= \left| -\frac{(\beta-1)\Delta}{2} - \Delta \right| \\ |e_Q(n)| &= \frac{(\beta+1)\Delta}{2} \end{aligned} \quad (4.12)$$

Substituindo $|e_Q(n)| = \frac{\Delta(\beta+1)}{2}$ e $|x(n)| = 1$ em (4.9) tem-se que o limite inferior de μ

considerando a potência de ruído, fica

$$\mu \geq \frac{1}{(\beta + 1)} \quad (4.13)$$

Da equação (4.10) tem-se que $\beta > \frac{2|z(n)|}{\Delta}$ e considerando $|z(n)|_{max} = 3\sigma$, então pode escrever-se que,

$$\beta = \frac{6\sigma}{\Delta} \quad \text{com } \beta \text{ ímpar}$$

Mas como σ é arbitrário e $\beta \in \mathbb{Z}$ definido ímpar, então escolhe-se o primeiro ímpar inteiro superior a $\frac{6\sigma}{\Delta}$

4.7.2 Limite superior do passo de adaptação em função de σ

Sabendo que para ter um estado absorvente na cadeia de Markov quando $w(n) = \hat{w}$, $y_Q(n)$ deve ser igual a zero para toda entrada $x(n)$, ou seja

$$|Q\{\mu e_Q(n)x(n)\}| < \Delta \quad (4.14)$$

ou equivalentemente,

$$|\mu e_Q(n)x(n)| < \frac{\Delta}{2} \quad (4.15)$$

em que

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{w(n)x(n)\}| \quad (4.16)$$

com $w(n) = \hat{w}$ o erro fica,

$$|e_Q(n)| = |Q\{\hat{w}x(n) + z(n)\} - Q\{\hat{w}x(n)\}| \quad (4.17)$$

Analisando a equação (4.17) tem-se que:

- Se $|x(n)| = 1$

$$|e_Q(n)| = |Q\{\hat{w} + z(n)\} - Q\{\hat{w}\}|$$

Para valores positivos e negativos de $z(n)$ o erro é o mesmo, observe a Figura 4.4.

Então, tem-se

$$\begin{aligned} |e_Q(n)| &= \left| \left(\hat{w} - \frac{(\beta - 1)\Delta}{2} \right) - \hat{w} \right| \\ |e_Q(n)| &= \frac{(\beta - 1)\Delta}{2} \end{aligned}$$

substituindo este resultado para $|x(n)| = 1$ na equação (4.15) tem-se que,

$$\mu < \frac{1}{\beta - 1} \quad (4.18)$$

Mas este resultado deve ser válido para todo $x(n)$, então analisando para $|x(n)| \neq 1$, tem-se

- Se $|x(n)| \neq 1$ também pode obter-se $e_Q(n) \neq 0$, por exemplo se $\hat{w}x(n) = \frac{k\Delta}{2}$ em que $k > 0$ e ímpar por ser um ponto de decisão na quantização (o resultado é aplicável para $k < 0$ pela simetria no quantizador para valores práticos de b). Assim,

$$|e_Q(n)| = \left| Q\left\{ \frac{k\Delta}{2} + z(n) \right\} - Q\left\{ \frac{k\Delta}{2} \right\} \right|$$

Sabendo que o pior caso para o estado do erro acontece para valores negativos de $z(n)$ (observe a Figura 4.4) e sendo k e β inteiros ímpar, tem-se

$$\begin{aligned} |e_Q(n)| &= \left| \left(\frac{k\Delta}{2} - \frac{\beta\Delta}{2} \right) - \frac{(k + 1)\Delta}{2} \right| \\ |e_Q(n)| &= \frac{(\beta + 1)\Delta}{2} \end{aligned}$$

Substituindo este valor em (4.15) e considerando que o pior caso para $|x(n)| \neq 1$

acontece quando $|x(n)| = 1 - \Delta$ (estado menor seguinte a $|x(n)| = 1$), então

$$\mu < \frac{1}{(\beta + 1)(1 - \Delta)} \quad (4.19)$$

Comparando este resultado com o obtido em (4.18), tem-se que a condição que melhor representa o limite superior do passo de adaptação para todo $x(n)$ é:

$$\mu < \frac{1}{(\beta + 1)(1 - \Delta)} \quad (4.20)$$

Assim, o passo de adaptação considerando a influência da potência de ruído está dado pela expressão,

$$\frac{1}{(\beta + 1)} \leq \mu < \frac{1}{(\beta + 1)(1 - \Delta)}$$

sabendo que $\Delta = 2^{1-b}$ também pode escrever-se,

$$\boxed{\frac{1}{\beta + 1} \leq \mu < \frac{2^b}{(\beta + 1)(2^b - 2)}} \quad (4.21)$$

em que $\beta < \frac{6\sigma}{\Delta}$ com β inteiro ímpar.

Observe que para o caso de ruído baixo $|z(n)| \ll \frac{\Delta}{2}$ teremos $\sigma \ll \frac{\Delta}{6}$. Sabendo que $\beta > \frac{6\sigma}{\Delta}$ teremos $\beta = 1$. Com $\beta = 1$, a expressão dada em (4.21) fica igual à condição para o passo de adaptação considerando ruído baixo determinada na equação (4.8).

4.7.3 Exemplos

Exemplo 1: Considerando para este exemplo $\hat{w} = 0,5$, $b = 3$, $\Delta = 0,25$ e $\sigma = 0,14$, tem-se que $\beta > 3,36$, portanto sabendo que β deve ser o primeiro inteiro ímpar então $\beta = 5$. Logo substituindo em (4.21) a condição de μ para ter uma cadeia quase-ergódica com um

único estado absorvente em $w(n) = \hat{w}$ é dada por:

$$0,167 \leq \mu < 0,222 \quad (4.22)$$

A seguir apresenta-se a matriz de probabilidades de transição P_w e as probabilidades de estado estacionário para quatro valores de μ , dentro e fora do intervalo determinado em (4.22).

Caso I: Para $\mu = 0,16$, valor abaixo o intervalo espera-se ter mais de um estado absorvente.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \left[\begin{array}{cccccccc} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{array} \right]$$

$$P_w = \left(\begin{array}{cccccccc} 0,87 & 0,13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,88 & 0,12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,89 & 0,11 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,99 & 0,01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,99 & 0,01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1,00 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \end{array} \right)$$

Efetivamente tem-se mais de um estado absorvente na matriz P_w . Observe que para $w(n) = 0,25, 0,5$ e $0,75$ a probabilidade de voltar ao mesmo estado é sempre 100%. Portanto, como observa-se a seguir as probabilidades estacionárias indicam que trata-se de uma cadeia não ergódica.

$$\begin{array}{c} \textbf{Estados} \\ \left[\begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} \right] \end{array} P_w^{600000} = \begin{array}{c} \left(\begin{array}{cccccccc} 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 & 0,00 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 1,00 \end{array} \right) \end{array}$$

Observe como as probabilidades de estado estacionário conseguem-se após 600000 passos para P_w , isto devido ao passo de adaptação muito pequeno o qual relaciona-se diretamente com a velocidade de convergência.

Caso 2: Para $\mu = 0,2$. Passo dentro do intervalo (4.22)

$$\begin{array}{c} \textbf{Estados} \\ \left[\begin{array}{c} -1,00 \\ -0,75 \\ -0,50 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} \right] \end{array} P_w = \begin{array}{c} \left(\begin{array}{cccccccc} 0,757 & 0,243 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,812 & 0,188 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,814 & 0,186 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,893 & 0,107 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,976 & 0,024 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,999 & 0,001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,001 & 0,999 \end{array} \right) \end{array}$$

A matriz P_w apresenta um único estado absorvente. No estado $w(n) = 0,5$ a probabilidade de voltar ao mesmo estado é 100%. A seguir as probabilidades estacionárias determinadas

em P_w^{40000} demonstram que trata-se de uma cadeia de Markov quase-ergódica com único estado absorvente.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w^{40000} = \begin{pmatrix}
 \begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} & \begin{array}{cccccccc}
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000 \\
 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 1,000 & 0,000
 \end{array}
 \end{pmatrix}$$

Observe como as probabilidades de estado estacionário são obtidos só após 40000 passos para P_w , demonstrando a direta relação entre o passo de adaptação e a velocidade de convergência.

Caso 3: Para $\mu = 0,25$ acima do intervalo, espera-se não ter estados absorventes.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w = \begin{pmatrix}
 \begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} & \begin{array}{cccccccc}
 0,600 & 0,400 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0,612 & 0,388 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0,635 & 0,365 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0,748 & 0,252 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0,888 & 0,112 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0,977 & 0,023 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0,001 & 0,999 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0,023 & 0,977
 \end{array}
 \end{pmatrix}$$

Tal como se esperava para um valor de passo μ acima do intervalo determinado em (4.22),

a matriz P_w não terá qualquer estado absorvente, a cadeia de Markov segue sendo quase-ergódica, observe P_w^{40000} , mas o desajuste do LMS é não nulo.

$$\begin{array}{c} \text{Estados} \end{array} \quad \left[\begin{array}{cccccccc} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{array} \right]$$

$$P_w^{40000} = \begin{pmatrix} 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 & 0,023 & 0,962 & 0,015 \end{pmatrix}$$

4.7.4 Passo de adaptação para valores práticos de bits ($b \geq 12$)

De (4.21) tem-se que para valores práticos de b ($b \geq 12$) a condição do passo de adaptação é

$$\mu \simeq \frac{1}{\beta + 1} \quad (4.23)$$

Assim, sabendo que $\beta > \frac{6\sigma}{\Delta}$ e substituindo $\Delta = 2^{(1-b)}$ tem-se,

$$\beta > \frac{6\sigma}{2^{(1-b)}}$$

ou melhor

$$\beta > 3\sigma 2^b$$

como $N = 2^b$, então

$$\beta > 3N\sigma$$

Para valores práticos de b tem-se que $\beta \cong 3N\sigma$. Logo, substituindo em μ , tem-se que

$$\mu \cong \frac{1}{3N\sigma}$$

Além disso, lembre que para o passo de adaptação dado por:

$$\mu < \frac{\Delta}{2} \quad (4.24)$$

a matriz P_w é igual à matriz identidade. Substituindo $\Delta = 2^{1-b}$ em (4.24) também pode escrever-se,

$$\mu < \frac{1}{N}$$

Então, com um passo $\mu < \frac{1}{N}$, $P_w = I$. Assim, a condição a satisfazer para não ter uma matriz $P_w = I$, para valores de bits práticos e considerando que \hat{w} é um dos níveis do quantizador é

$$\frac{1}{3N\sigma} > \frac{1}{N}$$

ou,

$$\sigma < \frac{1}{3}$$

Portanto conclui-se que para valores práticos de b , em que $\sigma \geq \frac{1}{3}$, não existe um passo de adaptação possível para ter só um estado absorvente na matriz de probabilidades P_w quando $w(n) = \hat{w}$.

4.8 CONCLUSÕES

Neste capítulo estudou-se as condições a serem satisfeitas por P_w para que o algoritmo LMS tenha um comportamento adequado quando implementado em precisão finita. Além disso apresentou-se uma análise para determinar os limites do passo de adaptação em função da potência de ruído. Importantes conclusões foram feitas para o algoritmo LMS operando em situações reais. Exemplos mostram a validade da condição do passo de adaptação proposto.

Capítulo 5

CONCLUSÕES

Neste trabalho foi desenvolvida uma nova modelagem do comportamento do algoritmo LMS quando implementado em precisão finita, aplicado à identificação de sistemas de transmissão digital. Os coeficientes do filtro adaptativo foram modelados como estados de uma cadeia de Markov de primeira ordem e a matriz de probabilidades de transição da cadeia foi determinada para o caso unidimensional. Sinais e coeficientes foram representados em precisão finita incluindo os efeitos de quantização e de saturação. Desta forma obtiveram-se informações importantes sobre o comportamento do algoritmo, que não são obtidas usando os modelos existentes na literatura.

Foi desenvolvido um modelo para a função de densidade de probabilidade (fdp) do coeficiente adaptativo, condicionada ao valor do coeficiente anterior e ao estado do sinal de entrada. Exemplos da matriz de probabilidades de transição P_w mostram que o modelo proposto descreve com bastante precisão o desempenho obtido pela simulação.

Foram determinadas as propriedades desejáveis de P_w para que o algoritmo LMS em precisão finita convirja para o ponto ótimo \hat{w} em regime permanente, independentemente do estado inicial de $w(n)$. Conhecidas essas propriedades determinou-se as condições a serem satisfeitas pelo passo de adaptação.

Inicialmente, determinou-se os limites do passo de adaptação considerando uma potência

de ruído muito baixa. Após, foi desenvolvida uma análise para determinar as condições do passo de adaptação em função da potência do ruído de medição.

Importantes resultados obtidos deste trabalho:

- Existe uma forte relação entre o passo de adaptação e a potência do ruído. Quanto maior a potência de ruído σ^2 , menor o valor máximo do passo μ , e portanto menor a velocidade de convergência.
- Para valores práticos de bits, assumindo uma potência de ruído baixa e o estado ótimo como um dos níveis de quantização, o passo de adaptação pode ser representado com a mesma precisão numérica dos sinais e os coeficientes, já que 0,5 sempre será um dos níveis de quantização para $b \geq 2$.
- Para valores práticos de bits e considerando \hat{w} um dos níveis de quantização dos coeficientes do filtro, tem-se que para $\sigma \geq \frac{1}{3}$ não existirá um passo de adaptação possível tal que a matriz P_w tenha apenas um estado absorvente quando $w(n) = \hat{w}$.

Resultados de simulação de Monte Carlo comprovam a validade das condições obtidas para o passo de adaptação.

5.1 PROPOSTAS PARA TRABALHOS FUTUROS

Algumas propostas para a continuação do trabalho podem ser sugeridas, entre elas:

- **Velocidade de convergência:** Utilizar ferramentas de cadeias de Markov para estabelecer uma relação entre o valor do passo de adaptação e a velocidade de convergência do algoritmo.
- **Desajuste em função do passo de adaptação:** Seria interessante determinar, do ponto de vista do projetista, o desajuste do algoritmo em função do passo de adaptação.

- **Análise para o caso multidimensional:** Estudar e modelar o comportamento do algoritmo LMS quando implementado em precisão finita para um número maior de coeficientes adaptativos (caso multidimensional), determinando os limites do passo de adaptação em função da potência do ruído.

Apêndice A

Função erf

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{y^2}{2}} dy$$

<i>x</i>	<i>erf(x)</i>	<i>x</i>	<i>erf(x)</i>	<i>x</i>	<i>erf(x)</i>
0,0500	0,0199	1,0500	0,3531	2,0500	0,4798
0,1000	0,0398	1,1000	0,3643	2,1000	0,4821
0,1500	0,0596	1,1500	0,3749	2,1500	0,4842
0,2000	0,0793	1,2000	0,3849	2,2000	0,4861
0,2500	0,0987	1,2500	0,3944	2,2500	0,4878
0,3000	0,1179	1,3000	0,4032	2,3000	0,4893
0,3500	0,1368	1,3500	0,4115	2,3500	0,4906
0,4000	0,1554	1,4000	0,4192	2,4000	0,4918
0,4500	0,1736	1,4500	0,4265	2,4500	0,4929
0,5000	0,1915	1,5000	0,4332	2,5000	0,4938
0,5500	0,2088	1,5500	0,4394	2,5500	0,4946
0,6000	0,2257	1,6000	0,4452	2,6000	0,4953
0,6500	0,2422	1,6500	0,4505	2,6500	0,4960
0,7000	0,2580	1,7000	0,4554	2,7000	0,4965
0,7500	0,2734	1,7500	0,4599	2,7500	0,4970
0,8000	0,2881	1,8000	0,4641	2,8000	0,4974
0,8500	0,3023	1,8500	0,4678	2,8500	0,4978
0,9000	0,3159	1,9000	0,4713	2,9000	0,4981
0,9500	0,3289	1,9500	0,4744	2,9500	0,4984
1,0000	0,3413	2,0000	0,4772	3,0000	0,4987

Apêndice B

Exemplos para \hat{w} diferentes dos estados de $w(n)$

A seguir apresentam-se alguns exemplos em que \hat{w} não pertence a um dos estados de $w(n)$. Considera-se para estes exemplos $\mu = 0,6$, $b = 3$ bits e uma potência de ruído baixa com $\sigma = 10^{-3}$. Para o primeiro caso considera-se o coeficiente ótimo como um dos níveis dos coeficientes $\hat{w} = 0,75$, no segundo caso $\hat{w} = 0,7$, no terceiro $\hat{w} = 0,65$ e assim até alcançar um valor de $\hat{w} = 0,55$. Observe como para o passo de adaptação, $\mu = 0,6$, o algoritmo converge para o estado mais próximo de \hat{w} . Para observar melhor esta situação apresenta-se a matriz de probabilidades de transição P_w e as probabilidades estacionárias.

B.1 Caso 1

Para $\hat{w} = 0,75$

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$
$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,500 & 0,375 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 \end{pmatrix}$$

Sendo \hat{w} um dos estados de $w(n)$ e estando o valor do passo de adaptação dentro do

intervalo determinado no Capítulo 4, a matriz P_w apresenta só um estado absorvente. P_w^{100} mostra que trata-se de uma cadeia quase-ergódica, em que o algoritmo LMS converge para o estado ótimo.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \end{pmatrix}$$

B.2 Caso 2

Para $\hat{w} = 0,7$

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,623 & 0,252 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,000 \end{pmatrix}$$

Observe como para um valor de \hat{w} bem perto do estado 0,75, a matriz P_w também apresenta só um estado absorvente e de acordo às probabilidades em P_w^{100} também trata-se de uma cadeia quase-ergódica, em que o algoritmo LMS converge para o estado ótimo.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1,00 \end{pmatrix}$$

B.3 Caso 3

Para $\hat{w} = 0,65$

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,625 & 0,251 & 0,124 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,876 & 0,124 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,001 & 0,999 \end{pmatrix}$$

A medida que \hat{w} começa a afastar-se do estado 0,75, o estado absorvente na matriz P_w desaparece, mas a probabilidade de convergir ao estado $w(n) = 0,75$ é alta. Observe de P_w^{100} que trata-se de uma cadeia quase-ergódica, em que o algoritmo LMS na média converge para o valor 0,7485.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{150} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,01 & 0,99 \end{pmatrix}$$

B.4 Caso 4

Para $\hat{w} = 0,625$

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0,25 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,625 & 0,312 & 0,063 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,937 & 0,063 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,063 & 0,937 \end{pmatrix}$$

Observe na matriz P_w como para \hat{w} igual à metade do intervalo de quantização, a probabilidade de cair no mesmo estado é a mesma para $w(n) = 0,5$ e $0,75$, portanto em estado estacionário (matriz P_w^{100}) existirá uma probabilidade de 50% de cair no estado $0,5$ e 50% de cair no estado $0,75$. Trata-se de uma cadeia quase-ergódica, em que o algoritmo LMS na média converge para o valor $0,625$.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,50 & 0,50 \end{pmatrix}$$

B.5 Caso 5

Para $\hat{w} = 0,6$

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,625 & 0,374 & 0,001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,999 & 0,001 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,124 & 0,876 \end{pmatrix}$$

A medida que o valor de \hat{w} começa a afastar-se da metade do intervalo de quantização e aproximar-se do estado 0,5, a probabilidade de cair neste estado, sendo $w(0) = 0,5$ é muito alta, observe na matriz P_w . As probabilidades estacionárias em P_w^{100} indicam que trata-se de uma cadeia quase-ergódica. O algoritmo LMS na média converge para o valor 0,5016.

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w^{100} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,99 & 0,01 \end{pmatrix}$$

B.6 Caso 6

Para $\hat{w} = 0,55$

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0,375 & 0,250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,375 & 0,375 & 0,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,500 & 0,125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,625 & 0,375 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,875 & 0,125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,125 & 0,875 \end{pmatrix}$$

Observe como para um valor bem perto do estado 0,5, a matriz P_w apresenta um estado absorvente e portanto as probabilidades estacionárias em P_w^{100} indicam que trata-se de uma cadeia quase-ergódica. O algoritmo LMS converge para o estado 0,5 com probabilidade de 100%.

$$\begin{array}{c} \textbf{Estados} \\ \left[\begin{array}{c} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{array} \right] \end{array} \quad \left[\begin{array}{cccccccc} -1 & -0,75 & -0,5 & -0,25 & 0 & 0,25 & 0,5 & 0,75 \end{array} \right]$$

$$P_w^{100} = \left(\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1,00 & 0 \end{array} \right)$$

Todos os casos anteriormente expostos validam os resultados obtidos sobre o passo de adaptação, ver Capítulo 4.

Apêndice C

Discussão do passo de adaptação para

$$\hat{w} \leq -1 + \Delta$$

Considerando que \hat{w} é um dos estados dos coeficientes e a potência de ruído é muito baixa, tem-se que os estados em que $\hat{w} \leq -1 + \Delta$ o passo de adaptação para que o algoritmo convirja para o coeficiente ótimo fica fora do intervalo determinado no Capítulo 4. Mas o anterior acontece quando o número de bits é pequeno. Observe a seguir.

C.1 Caso para $\hat{w} = -1$

C.1.1 Limite inferior do passo de adaptação

O limite inferior do passo de adaptação ao igual que no Capítulo 4 fica estabelecido pela condição que $y_Q(n)$ deve satisfazer para evitar a paralização do algoritmo quando $w(n) \neq \hat{w}$. Então com,

$$|y_Q(n)| \geq \Delta \text{ ou} \tag{C.1}$$

$$|\mu e_Q(n)x(n)| \geq \frac{\Delta}{2} \tag{C.2}$$

em que para um ruído muito baixo e com $\hat{w} = -1$ o erro quantizado é,

$$e_Q(n) = Q\{-x(n)\} - Q\{w(n)x(n)\}$$

Sabendo que o mínimo valor de erro acontece quando $w(n) = -1 + \Delta$ tem-se,

$$e_Q(n) = Q\{-x(n)\} - Q\{(-1 + \Delta)x(n)\}$$

- Se $x(n) = -1$ então o erro fica,

$$e_Q(n) = Q\{1\} - Q\{1 - \Delta\}$$

onde

$$e_Q(n) = 1 - \Delta - (1 - \Delta) = 0$$

com $e_Q(n) = 0$ não é possível estabelecer um passo de adaptação já que substituindo em (C.2) $\mu \geq \infty$. Então, analisando para o valor de $x(n)$ seguinte tem-se,

- Se $x(n) = -1 + \Delta$ então o erro fica

$$e_Q(n) = Q\{1 - \Delta\} - Q\{(1 - \Delta)^2\}$$

em que,

$$e_Q(n) = Q\{1 - \Delta\} - Q\{1 - 2\Delta + \Delta^2\}$$

desprezando Δ^2 tem-se,

$$e_Q(n) = 1 - \Delta - 1 + 2\Delta$$

então com $e_Q(n) = \Delta$ e substituindo na equação (C.2),

$$\mu \geq \frac{1}{2(1 - \Delta)} \quad (\text{C.3})$$

C.1.2 Limite superior do passo de adaptação

Para a determinação do limite superior do passo de adaptação considera-se a condição que deve satisfazer $y_Q(n)$ para conseguir um estado absorvente quando $w(n) = \hat{w}$, ou seja

$$|y_Q(n)| < \Delta \text{ ou} \quad (\text{C.4})$$

$$|\mu e_Q(n)x(n)| < \frac{\Delta}{2} \quad (\text{C.5})$$

em que para um ruído muito baixo, com $\hat{w} = -1$ e considerando que $w(n) = \hat{w}$ o erro quantizado fica,

$$e_Q(n) = Q\{-x(n)\} - Q\{-x(n)\}$$

Observe que independente do valor de $x(n)$ o erro quantizado sempre será $e_Q(n) = 0$ portanto aplicando este resultado na equação (C.5) tem-se que não existe um limite superior para μ quando $\hat{w} = -1$. Portanto a condição para conseguir um estado absorvente deve satisfazer que:

$$\mu \geq \frac{1}{2(1 - \Delta)}$$

Mas, observe que para valores de bits práticos, Δ é muito pequeno fazendo que a condição para o limite inferior do passo de adaptação seja a mesma que foi determinada no Capítulo 4.

C.2 Caso para $\hat{w} = -1 + \Delta$

C.2.1 Limite inferior do passo de adaptação

Seguendo a mesma análise que no caso para $\hat{w} = -1 + \Delta$ tem-se que o limite inferior do passo de adaptação é determinado pela condição que deve satisfazer $y_Q(n)$ para evitar a paralização do algoritmo quando $w(n) \neq \hat{w}$. Então com,

$$|y_Q(n)| \geq \Delta \text{ ou} \quad (\text{C.6})$$

$$|\mu e_Q(n)x(n)| \geq \frac{\Delta}{2} \quad (\text{C.7})$$

em que para um ruído muito baixo e com $\hat{w} = -1 + \Delta$ o erro quantizado é,

$$e_Q(n) = Q\{(-1 + \Delta)x(n)\} - Q\{w(n)x(n)\}$$

Sabendo que o mínimo valor de erro acontecerá quando $w(n)$ seja o estado seguinte ou o estado anterior de \hat{w} , então fazendo $w(n) = -1$ tem-se,

$$e_Q(n) = Q\{(-1 + \Delta)x(n)\} - Q\{-x(n)\}$$

- Se $x(n) = -1$ então o erro fica,

$$e_Q(n) = Q\{1 - \Delta\} - Q\{1\}$$

onde

$$e_Q(n) = 1 - \Delta - (1 - \Delta) = 0$$

com $e_Q(n) = 0$ não é possível estabelecer um passo de adaptação já que substituindo em (C.7) $\mu \geq \infty$. Então se faz a análise para o valor de $x(n)$ seguinte.

- Se $x(n) = -1 + \Delta$ então o erro fica

$$e_Q(n) = Q\{(1 - \Delta)^2\} - Q\{1 - \Delta\}$$

em que,

$$e_Q(n) = Q\{1 - 2\Delta + \Delta^2\} - Q\{1 - \Delta\}$$

desprezando Δ^2 tem-se,

$$e_Q(n) = 1 - 2\Delta - 1 + \Delta$$

então com $e_Q(n) = -\Delta$ e substituindo na equação (C.7),

$$\mu \geq \frac{1}{2(1 - \Delta)} \quad (\text{C.8})$$

C.2.2 Limite superior do passo de adaptação

Para a determinar o limite superior do passo de adaptação $y_Q(n)$ deve ser menor que um intervalo de quantização para conseguir um estado absorvente quando $w(n) = \hat{w}$, ou seja

$$\begin{aligned} |y_Q(n)| &< \Delta \text{ ou} \\ |\mu e_Q(n)x(n)| &< \frac{\Delta}{2} \end{aligned}$$

em que para um ruído muito baixo, com $\hat{w} = -1 + \Delta$ e considerando que $w(n) = \hat{w}$ o erro quantizado fica,

$$e_Q(n) = Q\{(-1 + \Delta)x(n) + z(n)\} - Q\{(-1 + \Delta)x(n)\}$$

- Se $x(n) = -1$ o valor do erro quantizado é $e_Q(n) = 0$ e não é possível estabelecer o limite superior para o passo de adaptação.
- Se $x(n) = -1 + \Delta$ tem-se que

$$e_Q(n) = Q\{(-1 + \Delta)^2 + z(n)\} - Q\{(-1 + \Delta)^2\}$$

em que desprezando o valor de Δ^2 tem-se

$$e_Q(n) = Q\{1 - 2\Delta + z(n)\} - Q\{-1 + 2\Delta\}$$

sendo $z(n)$ muito pequeno, o valor de erro quantizado é zero e com este resultado não se consegue determinar o limite superior do passo

- Se $x(n) = -1 + 2\Delta$ tem-se que

$$e_Q(n) = Q\{(-1 + \Delta)(-1 + 2\Delta) + z(n)\} - Q\{(-1 + \Delta)(-1 + 2\Delta)\}$$

ou pode escrever-se

$$e_Q(n) = Q\{\frac{k\Delta}{2} + z(n)\} - Q\{\frac{k\Delta}{2}\}$$

em que $\frac{k\Delta}{2}$ é a metade do intervalo de quantização (ponto crítico), então o mínimo erro que pode alcançar é $e_Q(n) = \Delta$. Logo, substituindo os resultados na equação (C.5) tem-se que,

$$\mu < \frac{1}{2(1 - 2\Delta)}$$

Portanto, a condição do passo de adaptação para conseguir um estado absorvente quando $\hat{w} = -1 + \Delta$ é:

$$\frac{1}{2(1 - \Delta)} \leq \mu < \frac{1}{2(1 - 2\Delta)}$$

Mas observe que para valores de bits práticos o limite do passo de adaptação será $\mu \cong 0,5$, validando os resultados do Capítulo 4.

C.3 Exemplos

A seguir apresentam-se alguns exemplos para $b = 3$ bits, $\sigma = 10^{-3}$ e limites para o passo de adaptação de $0,667 \leq \mu < 1$ se $\hat{w} = -1 + \Delta$ e $\mu \geq 0,667$ se $\hat{w} = -1$.

C.3.1 Caso 1

A matriz P_w para $\hat{w} = -1$ com um passo $\mu = 50$ é:

$$\begin{array}{c} \textbf{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$P_w = \begin{pmatrix} 1,000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,250 & 0,750 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,625 & 0 & 0,375 & 0 & 0 & 0 & 0 & 0 \\ 0,875 & 0 & 0 & 0,125 & 0 & 0 & 0 & 0 \\ 0,875 & 0 & 0 & 0 & 0,125 & 0 & 0 & 0 \\ 0 & 0,875 & 0 & 0 & 0 & 0,125 & 0 & 0 \\ 0 & 0 & 0,875 & 0 & 0 & 0 & 0,125 & 0 \\ 0 & 0 & 0 & 0,875 & 0 & 0 & 0 & 0,125 \end{pmatrix}$$

Observe a presença de um estado absorvente quando $w(n) = \hat{w}$. O passo de adaptação é um valor alto, lembre que não existe limite superior de μ quando $\hat{w} = -1$. A cadeia de Markov segue sendo quase-ergódica.

C.3.2 Caso 2

A matriz P_w para $\hat{w} = -1 + \Delta$ em que $\mu = 0,5$ está dada por:

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 1,000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1,000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,125 & 0,875 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,125 & 0,875 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,125 & 0,375 & 0,500 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,125 & 0,500 & 0,375 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,250 & 0,375 & 0,375 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,250 & 0,375 & 0,375 \end{pmatrix}$$

Neste caso o valor do passo de adaptação é menor que o limite inferior, isso justifica a presença de mais de um estado absorvente na matriz P_w , observe os estados -1 e $-0,75$. A cadeia de Markov segue sendo quase-ergódica

C.3.3 Caso 3

A seguir apresenta-se a matriz P_w quando $\hat{w} = -1 + \Delta$ e $\mu = 0,8$

$$\begin{array}{c} \text{Estados} \end{array} \quad \begin{bmatrix} -1,00 & -0,75 & -0,50 & -0,25 & 0,000 & 0,250 & 0,500 & 0,750 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ -0,75 \\ -0,5 \\ -0,25 \\ 0 \\ 0,25 \\ 0,5 \\ 0,75 \end{bmatrix} P_w = \begin{pmatrix} 0,750 & 0,250 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1,000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,125 & 0,875 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,125 & 0,250 & 0,625 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,125 & 0,375 & 0,500 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,375 & 0,250 & 0,375 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,375 & 0,250 & 0,375 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,438 & 0,187 & 0,375 \end{pmatrix}$$

Estando $\mu = 0,8$ dentro dos limites estabelecidos quando $b = 3$, a matriz P_w apresenta um único estado absorvente quando $w(n) = \hat{w}$

Apêndice D

Exemplos para o caso multidimensional

A seguir apresentam-se alguns exemplos para entrada e coeficientes vetoriais (caso multidimensional). O propósito destes exemplos é ilustrar que as propriedades do algoritmo LMS em precisão finita obtidas nesta dissertação para o caso escalar, podem ser aplicáveis ao caso multidimensional. Observe a seguir como para um determinado passo de adaptação ($\mu = 0,27$) o algoritmo sempre termina convergindo para \hat{w} independentemente do valor inicial de $w(n)$, característica de uma classe ergódica com um único estado absorvente. No entanto para outro valor de passo de adaptação observa-se uma dependência do estado inicial.

Muitos testes e mais análise devem ser realizados para assegurar que as propriedades aqui expostas também se cumpram para caso multidimensional.

D.1 Exemplo 1

Considera-se para este exemplo $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $b = 3$, $\sigma = 10^{-3}$ e média de 1000 realizações. Apresentam-se quatro casos em que $\mu = 0,27$ para diferentes estados iniciais $w(0)$.

D.1.1 Caso 1:

Para $\mu = 0,27$ e $w(0) = [0; 0; 0; 0]$

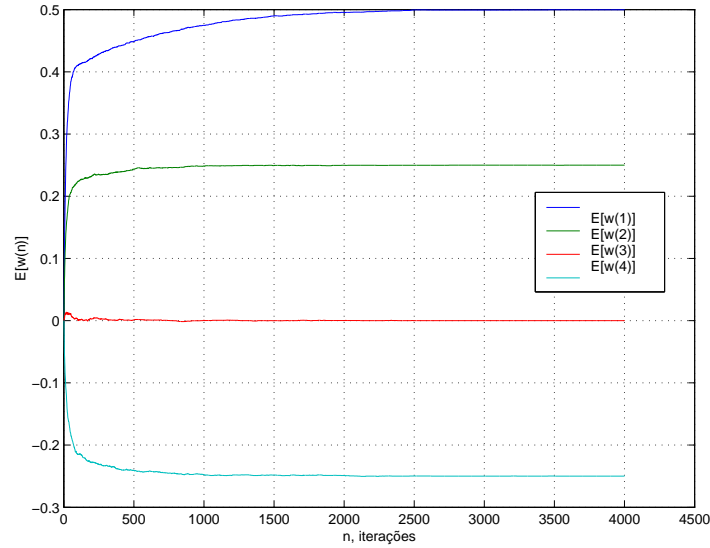


Figura D.1: Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0; 0; 0; 0]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$

D.1.2 Caso 2:

Para $\mu = 0,27$ e $w(0) = [-0,50; -0,25; 0,00; 0,25]$

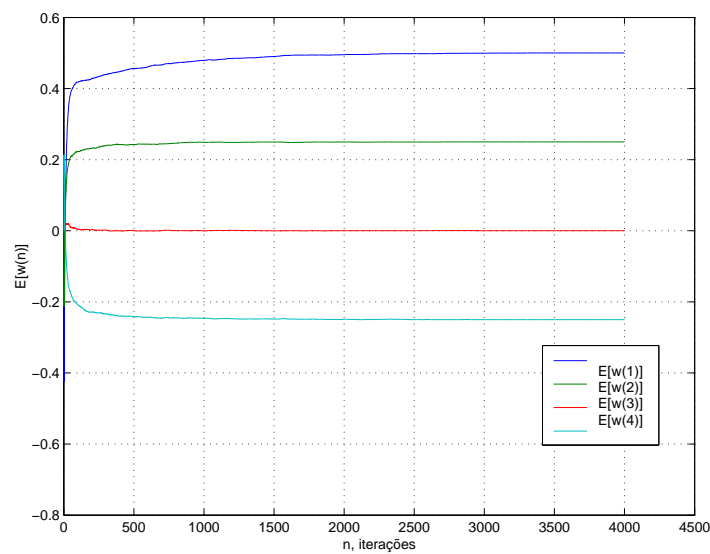


Figura D.2: Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-0,50; -0,25; 0,00; 0,25]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$

D.1.3 Caso 3:

Para $\mu = 0,27$ e $w(0) = [-1; -1; -1; -1]$

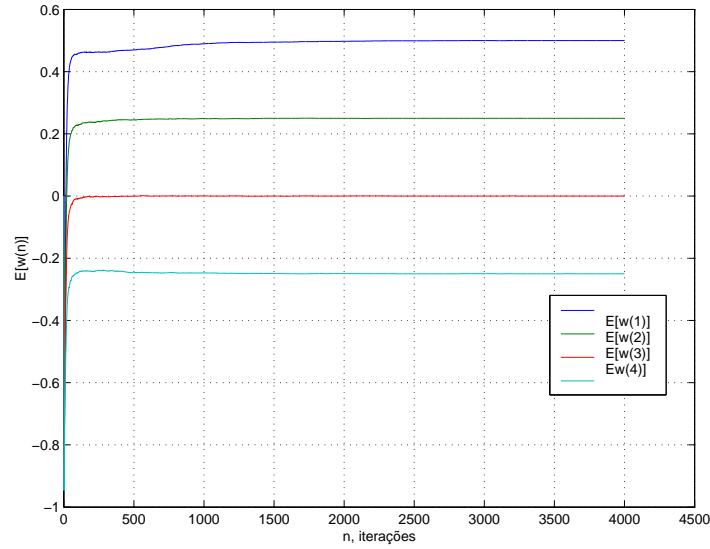


Figura D.3: Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-1; -1; -1; -1]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$

D.1.4 Caso 4:

Para $\mu = 0,27$ e $w(0) = [0,25; 0,5; -0,25; 0]$

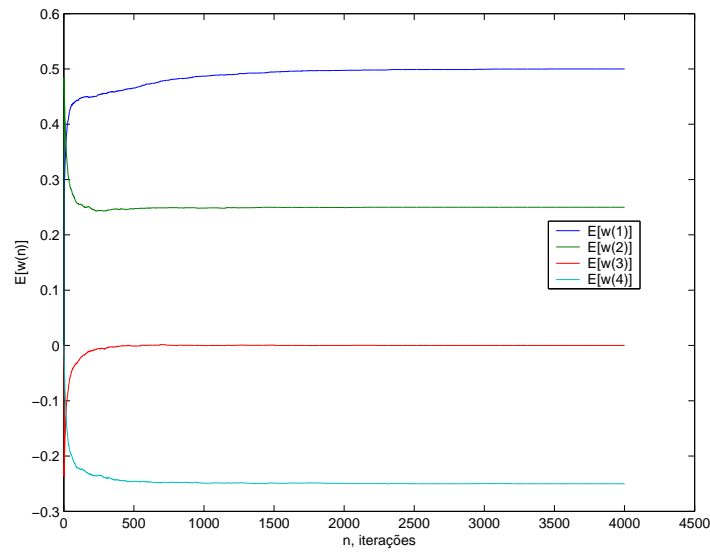


Figura D.4: Comportamento algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0,25; 0,5; -0,25; 0]$, $\mu = 0,27$, 1000 realizações e $\sigma_z = 10^{-3}$

Observe como em todos estes casos, para um passo de adaptação $\mu = 0,27$, ainda que com diferentes estados iniciais $w(0)$, o algoritmo converge para o estado ótimo $\hat{w} = [-0,50; -0,25; 0,00; 0,25]$. Isto indica a presença de estados absorventes, situação semelhante ao caso escalar.

D.2 Exemplo 2

Considera-se para este exemplo $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $b = 3$, $\sigma = 10^{-3}$ e média de 1000 realizações. Apresentam-se quatro casos em que $\mu = 0,2$ para diferentes estados iniciais $w(0)$.

D.2.1 Caso 1:

Para $\mu = 0,2$ e $w(0) = [0; 0; 0; 0]$

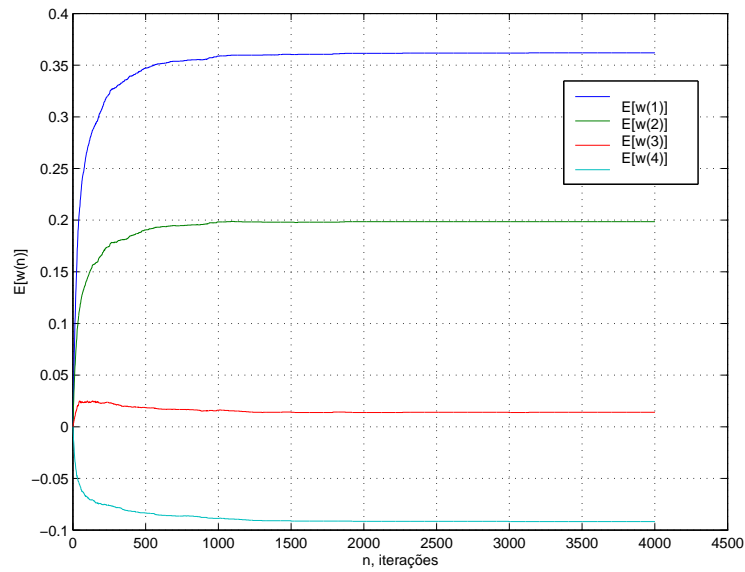


Figura D.5: Comportamento do algoritmo LMS para $b = 3\text{bits}$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0; 0; 0; 0]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$

D.2.2 Caso 2:

Para $\mu = 0,2$ e $w(0) = [-0,50; -0,25; 0,00; 0,25]$

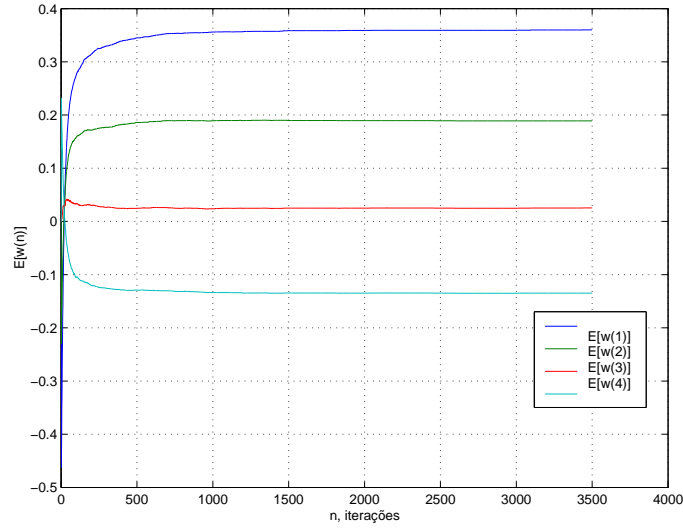


Figura D.6: Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-0,50; -0,25; 0,00; 0,25]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$

D.2.3 Caso 3:

Para $\mu = 0,2$ e $w(0) = [-1; -1; -1; -1]$

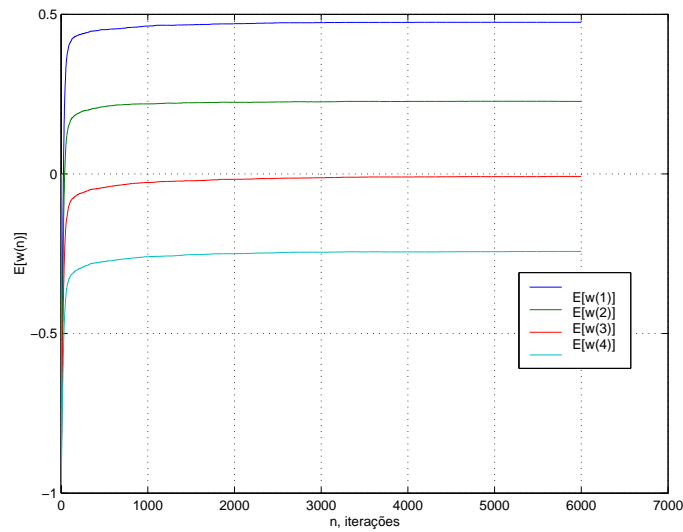


Figura D.7: Comportamento do algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [-1; -1; -1; -1]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$

D.2.4 Caso 4:

Para $\mu = 0,2$ e $w(0) = [0,25; 0,5; -0,25; 0,00]$

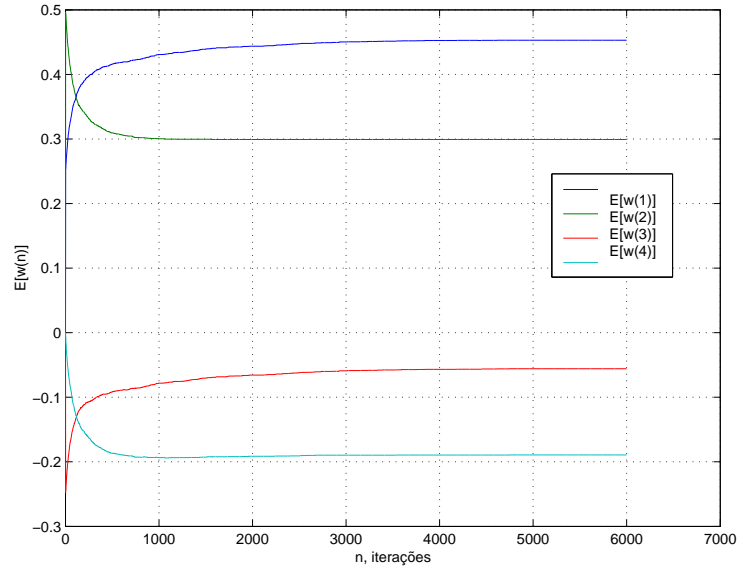


Figura D.8: Comportamento algoritmo LMS para $b = 3bits$, $\hat{w} = [0,50; 0,25; 0,00; -0,25]$, $w(0) = [0,25; 0,5; -0,25; 0,00]$, $\mu = 0,2$, 1000 realizações e $\sigma_z = 10^{-3}$

Observe que para um valor de passo $\mu = 0,2$ o algoritmo converge para coeficientes diferentes dependendo do estado inicial de $w(n)$.

Apêndice E

Símbolos e Abreviaturas

- $d(n)$ - sinal desejado em precisão infinita.
- $d_Q(n)$ - sinal desejado quantizado.
- $e(n)$ - sinal de erro não saturado.
- $e_Q(n)$ - sinal de erro quantizado.
- EMQ - erro médio quadrático.
- fdp - função densidade de probabilidade.
- FIR - finite impulse response.
- IIR - infinite impulse response.
- N - número de níveis do quantizador.
- n - tempo discreto ou número da iteração no algoritmo adaptativo.
- $w(n)$ - coeficiente do algoritmo adaptativo na n iteração.
- \hat{w} - coeficiente ótimo.
- $x(n)$ - sinal de entrada ou sinal de referência em precisão finita.
- $\hat{d}(n)$ - sinal da saída do filtro adaptativo.
- $z(n)$ - ruído de medição.

Símbolos especiais:

- μ - passo de adaptação.

- σ^2 - variância do sinal do ruído aditivo.
- **Q** - Quantizador.
- P_w - Matriz de probabilidades de transição de estados de uma cadeia de Markov.

Referências Bibliográficas

- [1] S.Haykin, *Adaptive Filter Theory*, Fourth edition, Prentice–Hall, Inc.,New Jersey, 2002
- [2] D.G. Manolakis, V.K. Ingle and S.M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering and array processing*, McGraw–Hill, USA, 2000.
- [3] J.C.M. Bermudez e N.J. Bershad, *A nonlinear analytical model for the quantized LMS algorithm - The arbitrary step size case*, IEEE Trans. Signal Processing, vol.44, no. 5, pp. 1175-1183, May 1996.
- [4] C.Caraiscos e B.Liu, *A roundoff error analysis of LMS adaptive algorithm*, IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-32, no. 1, pp. 34-41, Feb.1984.
- [5] S.T. Alexander, *Transient weighth misadjustment properties for the finite precision LMS algorithm*, IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-35, no. 9, pp. 1250-1258, Sept.1984.
- [6] J.M. Cioffi, *Limited precision effects in adaptive filtering*, IEEE Trans. on circuits and systems, vol.CAS-34, no. 7, pp. 821-833, July 1987.
- [7] J.C.M. Bermudez e N.J. Bershad, *Transient and tracking performance analysis of the quantized LMS algorithm for time-varing system identification*, IEEE Trans. Signal Processing, vol.44, no. 8, pp. 1990-1997, August 1996.
- [8] J.C.M. Bermudez e N.J. Bershad, *New insights on the transient and steady-state behavior of the quantized LMS algorithm*, IEEE Trans. Signal Processing, vol.44, no. 10, pp. 2623-2625, October 1996.
- [9] J.C.M. Bermudez e N.J. Bershad, *A nonlinear analytical model for the quantized LMS algorithm - The power-of-two step size case*, IEEE Trans. Signal Processing, vol.44, no. 11, pp. 2895-2900, November 1996.
- [10] J.Salazar, *Procesadores Digitales de Señal (DSP). Arquitectura y criterios de selección*, Universidad politécnica de cataluña.
- [11] A. V. Oppenheim, A.S. Willsky e S.H. Nawad, *Signals and Systems*, Prentice Hall, New Jersey, 1996.
- [12] R. Gupta e A.O. Hero, *IEEE Power versus Performance Tradeoffs for Reduced Resolution LMS Adaptive Filters*, IEEE Trans.Signal Processing, vol. 48, no.10, October 2000.

- [13] P.Z. Peebles, Jr, *Probability, random variables, and random signal principles*, third edition, McGraw–Hill, USA, 1993.
- [14] P.Brémaud, *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues*, Springer, New York, 1999.
- [15] J.B. Fonollosa, J.M. Sallán, A. Suñé, *Métodos cuantitativos de organización industrial II*, Ediciones UPC, Barcelona, 2002
- [16] J.R.Norris, *Markov Chains*, Cambridge, New York, 1997.
- [17] H. Besbes, M. Jaidane–Saidane, J. Ezzine, *On exact convergence results of adaptive filters: the finite alphabet case*, Elsevier, Signal Processing, 2000.